(2)

RADC-TR-84-72
Final Technical Report
April 1984

ALGORITHMS FOR RECONSTRUCTION OF PARTIALLY KNOWN, BAND LIMITED FOURIER TRANSFORM PAIRS FROM NOISY DATA

RGB Associates, Inc.

Sponsored by
Defense Advanced Research Projects Agency (DOD)
ARPA Order No. 3655

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

ROME AIR DEVELOPMENT CENTER
Air Force Systems Command
Griffiss Air Force Base, NY 13441

This report has been reviewed by the RADC Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

RADC-TR-84-72 has been reviewed and is approved for publication.

APPROVED: R J Michalak

RICHARD J. MICHALAK
Project Engineer

APPROVED: Lawrence Hillebrand

LAWRENCE J. HILLEBRAND, Colonel, USAF
Chief, Surveillance Division

FOR THE COMMANDER: John A. Ritz

JOHN A. RITZ
Acting Chief, Plans Office

If your address has changed or if you wish to be removed from the RADC mailing list, or if the addressee is no longer employed by your organization, please notify RADC (OCSE) Griffiss AFB NY 13441. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document requires that it be returned.

ALGORITHMS FOR RECONSTRUCTION OF PARTIALLY KNOWN, BAND LIMITED FOURIER TRANSFORM PAIRS FROM NOISY DATA

Richard Barakat

Contractor: RGB Associates, Inc.
Contract Number: F30602-82-C-0146
Effective Date of Contract: 4 August 1982
Contract Expiration Date: 30 April 1983
Short Title of Work: Phase Deconvolution Studies
Program Code Number: 3E20
Period of Work Covered: Aug 82 - 30 Apr 83

Principal Investigator: Richard Barakat
(617) 358-4675

Project Engineer: Dr. Richard J. Michalak
(315) 330-3143

This research was supported by the Defense Advanced Research Projects Agency of the Department of Defense and was monitored by Dr. R.J. Michalak (OCSE), Griffiss AFB NY 13441 under Contract F30602-82-C-0146

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION: UNCLASSIFIED	1b. RESTRICTIVE MARKINGS: N/A
2a. SECURITY CLASSIFICATION AUTHORITY: N/A	3. DISTRIBUTION/AVAILABILITY OF REPORT: Approved for public release; distribution limited. UN
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE: N/A	
4. PERFORMING ORGANIZATION REPORT NUMBER(S): N/A	5. MONITORING ORGANIZATION REPORT NUMBER(S): RADC-TR-84-72
6a. NAME OF PERFORMING ORGANIZATION: RGB Associates, Inc. / 6b. OFFICE SYMBOL (If applicable):	7a. NAME OF MONITORING ORGANIZATION: Rome Air Development Center (OCSE)
6c. ADDRESS (City, State and ZIP Code): P O Box 8, Wayland MA 01778	7b. ADDRESS (City, State and ZIP Code): Griffiss AFB NY 13441
8a. NAME OF FUNDING/SPONSORING ORGANIZATION: Defense Advanced Research Projects Agency / 8b. OFFICE SYMBOL (If applicable): STO	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: F30602-82-C-0146
8c. ADDRESS (City, State and ZIP Code): 1400 Wilson Blvd, Arlington VA 22209	10. SOURCE OF FUNDING NOS. — PROGRAM ELEMENT NO.: 62301E; PROJECT NO.: C655; TASK NO.: 01; WORK UNIT NO.: 08

11. TITLE (Include Security Classification): ALGORITHMS FOR RECONSTRUCTION OF PARTIALLY KNOWN, BAND LIMITED FOURIER TRANSFORM (Cont'd)

12. PERSONAL AUTHOR(S): Richard Barakat

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
Final	FROM Aug 82 TO Apr 83	April 1984	236

16. SUPPLEMENTARY NOTATION: None
None

17. COSATI CODES		
FIELD	GROUP	SUB. GR.

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number): Fourier Transform Pairs, Singular Value Decomposition, Phase Retrieval, Wave-front Aberrations.

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This report summarizes detailed mathematical investigations, both analytical and numerical, on recovery of partially known Fourier transforms with emphasis in the context of optical diffraction theory.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT: UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION: UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE NUMBER (Include Area Code)	22c. OFFICE SYMBOL
Dr. Richard J. Michalak	(315) 330-3143	RADC (OCSE)

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

Block 11 (Continued)

PAIRS FROM NOISY DATA

GENERAL INTRODUCTION

A recurring problem in many fields, especially diffraction optics, is the reconstruction of a Fourier transform pair g,G from partial data on either or both functions. Considerable effort has been expended in the development of algorithms for its solution; although there have been some successes, the problem has generally proved to be difficult. Of particular importance to the RADC effort is the retrieval of wavefront aberrations from the measured point spread function of an optical system.

Discussions with several investigators who have employed their own algorithms to these problems have indicated a sort of hit-or-miss attitude with respect to their behavior in various situations. Sometimes the particular algorithm works and sometimes it fails when the data are noisy. With the possible exception of Youla's recent study, there are really no serious attempts to understand the stability, rate of convergence, etc. with respect to noise in the measurements.

The present contract effort was devoted to the development and mathematical understanding of new algorithms based upon numerical functional analysis which are robust with respect to noisy data. The basic material is contained in the two sections entitled:

1. *Algorithms for reconstruction of partially known, bandlimited Fourier transform pairs from noisy data: I, the prototypical linear problem.*

II. *Algorithms for reconstruction of partially known, bandlimited Fourier transform pairs from noisy data: II, the nonlinear problem of phase retrieval.*

Both sections are very mathematical and employ mathematics not commonly encountered by optical physicist and engineers. For this reason a summarizing section has been included; it is the first section in the report and is entitled

Algorithms for reconstruction of partially known, bandlimited Fourier transform pairs from noisy data.

ALGORITHMS FOR RECONSTRUCTION OF PARTIALLY KNOWN, BANDLIMITED FOURIER TRANSFORM PAIRS FROM NOISY DATA

ABSTRACT

This paper is a summary of more detailed mathematical work by the author on recovery of partially known Fourier transforms. These problems of inversion of the finite Fourier transform and of phase retrieval are known to be ill-posed. We draw a distinction in the resultant ill-conditioning of the problems between global ill-conditioning (due to the existence of multiple exact solutions) and local ill-conditioning (due to the existence of large neighborhoods of the true solution, all of whose members are indistinguishable from the true solution if the data is noisy). We then develop extensions of known algorithms that attempt to reduce at least the effects of local ill-conditioning on numerical solutions by using the idea of filtered singular value decomposition, and present some numerical examples of the use of those algorithms. The originator-supplied key words include: → to p. 473

1. INTRODUCTION

A recurring problem in many fields (especially diffraction optics, electron microscopy, and X-ray diffraction) is the reconstruction of a Fourier transform pair g,G from partial data on either or both functions. Considerable effort has been expended in the development of algorithms for its solution; although there have been some successes, the problem has generally proved to be difficult.

The canonical examples for such reconstructions are:

Example 1, Extrapolation of Band-Limited Signals. Given a noisy measurement $\tilde{g}$ of g on an interval $A \equiv [a_1,a_2]$ and the knowledge that G vanishes outside the bounded interval $B \equiv [b_1,b_2]$, reconstruct g and G on the entire real line.

Example 1 is the archetypical linear problem in transform reconstruction. A number of algorithms have been proposed for its solution, either by iterative means: Gerschberg and Saxton [1], Papoulis [2], Youla [3] or by direct means: Cadzow [4], Sabri and Steenaart [5].

Example 2, The Phase Problem. Given a noisy measurement m of $|g|$ on an interval A and the knowledge that G vanishes outside the bounded interval B, reconstruct g and G on the entire real line.

Example 2 has been of theoretical interest for some time; see for example Burge, Fiddy, Greenway, and Ross [6]; however in this most general form it has proved intractable. Numerical solutions obtained in particular cases have done so by (sensibly) incorporating further knowledge of g and G. Two

such examples are:

Example 2a, The Two Moduli Problem. Given noisy measurements m and n of $|g|$ on A and $|G|$ on B, and the knowledge that G vanishes identically outside of B, reconstruct g,G over the entire real line.

Example 2b, The Phase Problem with Nonnegativity Constraints. Given a noisy measurement m of $|g|$ on A and the knowledge that G is nonnegative over B and vanishes identically outside B reconstruct g,G over the real line.

Gerschberg and Saxton [1] developed a widely used algorithm for Example 2a; almost all the iterative algorithms for the solution of the general reconstruction problem are suitably modified versions of this particular case. The Gerschberg-Saxton algorithm (hereafter denoted as the GS algorithm) in its general form is essentially the steepest descent algorithm with unit step length and so is first order [7]. In [8], the GS algorithm is successfully applied to a problem of the type in Example 2b. An alternative second order algorithm, based on Newton's method, has been proposed by Barakat and Newsam [9].

The canonical examples presented are simple in form, nevertheless they contain the salient features that make other, more complicated, problems intractable. The font of all difficulties is that the reconstruction problem is ill-posed, and badly ill-posed at that. The original definition of a well posed problem is due to Hadamard [10].

Definition. A problem is well posed if the solution

1. exists
2. is unique
3. depends continuously on the data.

If a problem violates any of these conditions, it is ill-posed. Each of the model problems violates at least one of these criteria. The literature, e.g. [11-13] has focussed on violations of uniqueness, however it is our contention that violation of condition 3 is the cause of the most of the problems encountered in numerical solutions; in particular it accounts for the extreme sensitivity of such solutions to small perturbations in the data.

The purpose of this paper is to summarize for the optical community the detailed analysis of transform reconstructions developed in [14-16]. These three references contain a detailed mathematical treatment of ill-conditioning in transform recovery problems with emphasis on the implications for numerical algorithms. The theory is *dimension independent*, although the supporting numerical calculations are restricted to one-dimensional problems. We hope to present two-dimensional calculations in the near future.

In order to present the results the following notation will be used. If $D \subset \mathbb{R}^N$ then $L^2(D)$ denotes the space of square integrable, complex valued functions over D with norm $\|\cdot\|$ and inner product $(\cdot,\cdot)$. If T is any set in $L^2(\mathbb{R}^N)$ then the projection P_T onto T is defined by

$$y = P_T x \iff y \in T \quad \text{and} \quad \|y - x\| = \inf_{z \in T} \|z - x\| \tag{1.1}$$

In the special case when $T \equiv L^2(D)$; $P_{L^2(D)}$ will be abbreviated to P_D; P_D has the form

$$(P_D G)(\hat{\omega}) = \begin{cases} G(\hat{\omega}) & \hat{\omega} \in D \\ 0 & \text{otherwise} \end{cases} . \tag{1.2}$$

The Fourier transform $\mathcal{F}: L^2(\mathbb{R}^N) \to L^2(\mathbb{R}^N)$ is defined to be

$$g(\hat{v}) = (\mathcal{F}G)(\hat{\omega}) = \int_{-\infty}^{\infty} e^{2\pi i \hat{v}\cdot\hat{\omega}} G(\hat{\omega})\, d\hat{\omega} . \tag{1.3}$$

If D and E are bounded subsets of $\mathbb{R}^N$ then the operator $P_D \mathcal{F} P_E$ will be called a finite Fourier transform (fFT). Finally the interval $[-c,c]$ shall be denoted by cI; the projection P_{cI} shall be abbreviated to P_c, and the fFT $P_c \mathcal{F} P_c$ is represented by $\mathcal{F}_c$.

The next section of the paper presents a survey of known results on example 1, which may be shown to be equivalent to inversion of the operator $\mathcal{F}_c$ where $c = \frac{1}{2}\pi[(a_1-a_2)(b_1-b_2)]^{1/2}$. Of particular importance is the relation between the ill-conditioning of problem 1 and the singular value decomposition (SVD) of $\mathcal{F}_c$; and how the two ideas are combined in inversion by filtered SVD. In Section 3 local ill-conditioning of the nonlinear phase retrieval problems is described in terms of that of the fFT; and it is contrasted with the global ill-conditioning due to the possibility of multiple exact solutions. Section 4 presents a simple generalization of the GS algorithm, tailored to overcome some of this ill-conditioning; together with a review of its convergence properties. In Section 5 extensions of this

algorithm are discussed, Section 6 contains a *brief* survey of numerical results (in the context of optical diffraction theory) on the relative behavior of these algorithms, followed by some concluding remarks.

2. THE LINEAR PROBLEM

The extrapolation problem of example 1 requires for its solution inversion of the linear integral equation

$$P_A \mathcal{F} P_B G = P_A g \quad . \tag{2.1}$$

Since A and B are bounded sets $P_A \mathcal{F} P_B$ is a compact linear integral operator and Eq. (2.1) is a Fredholm integral equation of the first kind. Inversion of such an equation is the prototypical linear ill-posed problem.

The ill-posed nature of the problem is exposed in the construction of the solution using the singular value decomposition (SVD) of $P_A \mathcal{F} P_B$. As outlined in Baker [17], a compact linear operator has an SVD $\{\phi_i, \sigma_i, \psi_i\}_{i=1}^{\infty}$ consisting of functions ϕ_i and ψ_i and nonnegative real numbers σ_i with the properties that

i. $\{\phi_i\}_{i=1}^{\infty}$ and $\{\psi_i\}_{i=1}^{\infty}$ are complete sets of orthonormal functions for $L^2(A)$ and $L^2(B)$, respectively.

ii. $\sigma_i \geq \sigma_{i+1}$ and $\lim_i \sigma_i = 0$.

iii. $P_A \mathcal{F} P_B \psi_i = \sigma_i \phi_i$. (2.2)

Expansion of G and $P_A g$ as sums of singular functions

$$G = \sum_i b_i \psi_i \qquad P_A g = \sum_i a_i \phi_i \tag{2.3}$$

gives a formal solution to Eq. (2.1) by equating coefficients, i.e.,

$$P_A \mathcal{F} P_B G = P_A g \iff \sigma_i b_i = a_i \quad . \tag{2.4}$$

This solution illustrates the ill-conditioning in the problem. For instance if $\tilde{g}$ is a perturbation of g such that

$$P_A\tilde{g} = \sum_i \tilde{a}_i\phi_i \qquad \text{with} \qquad \sum_i |\tilde{a}_i - a_i|^2 < \varepsilon^2 \tag{2.5}$$

then it is possible that the error occurs in a high frequency (large i) component of $P_A g$, so that a perturbation of size $\varepsilon\sigma_i^{-1}$ is induced in G. For fixed ε this perturbation grows arbitrarily large as $i\to\infty$. Thus Eq. (2.1) fails condition 3 of Hadamard's definition with respect to data perturbation in g. A similar argument, but one that is rarely made, shows that the equation is also ill-posed with respect to changes in the model, i.e. perturbations of the operator $P_A\mathcal{F}P_B$.

The above arguments show that the whole solution G cannot be recovered to within any specified accuracy given uncertainty in the data. Therefore the natural question to ask next is: "How large a part of the solution can be recovered to within a desired accuracy in the presence of noise?" A partial answer lies in the idea of the *essential dimension* $N(\delta,\varepsilon_1,\varepsilon_2)$ of the problem; loosely speaking this is the maximum number of parameters in a description of the solution that can be determined to within an accuracy δ, given errors ε_1 in $P_A g$ and ε_2 in $P_A\mathcal{F}P_B$. (The latter error includes discretization and roundoff errors as well as those arising from an imperfect mathematical model of the real world.)

A more precise definition of the essential dimension as the maximum dimension of any subspace U for which the associated projection $P_U G$ of the solution can be accurately calculated [16,18] shows that it may be

expressed in terms of the singular values. The optimum subspace is the span of the first $N(\delta,\varepsilon_1,\varepsilon_2)$ singular functions. This analysis suggests that Eq. (2.1) be solved by filtered SVD. This algorithm was introduced by Hanson [19] in which G is expressed as the sum

$$G = \sum_i p(a_i,\sigma_i,\gamma)\psi_i \tag{2.6}$$

where p is a filter function with the general form

$$\begin{aligned} \mathcal{P}(a,\sigma,\gamma) &\sim \sigma^{-1} \qquad \text{for large } \sigma \\ &\sim 0 \qquad \text{for small } \sigma \end{aligned} \tag{2.7}$$

and γ is a parameter that incorporates knowledge of data errors and the desired solution accuracy.

Filters come in many forms. For instance, if the error in g is such that $\|P_A g - P_A \tilde{g}\| \leq \varepsilon_1$, the model error ε_2 is negligible and the projection $P_U G$ is to be determined to within an accuracy δ (i.e. $\|P_U G - P_U \tilde{G}\| \leq \delta$); then the filter associated with the essential dimension is

$$\begin{aligned} p(a,\sigma,\gamma) &= a\sigma^{-1} \quad \text{if} \quad \sigma \geq \gamma \\ &= 0 \qquad\qquad\quad \sigma < \gamma \end{aligned} \tag{2.8}$$

where $\gamma = \varepsilon_1 \delta^{-1}$. This cutoff filter is a special case $(q \to \infty)$ of the class of filters

$$p(a,\sigma,\gamma) = \frac{\sigma^q}{\sigma^{q+1} + \gamma^{q+1}} \quad . \tag{2.9}$$

The utility of the theory of the essential dimension lies in its ability to predict the size and general form of components of the solution that may be

accurately recovered in the presence of noise, before numerical calculations are undertaken. Such predictions are therefore most useful in deciding on the size and form of an appropriate discretization for use in numerical solutions. The theory is easily applied to transform recovery problems due to the large body of knowledge about the SVD of the finite Fourier transform (fFT) developed by Landau, Pollak, Slepian and Walom [20-26]. The next theorem gives a brief summary of those results that are most useful in the present problem.

Theorem 1. i. If D and E are bounded subsets of $\mathbb{R}^N$ with volumes $|D|$ and $|E|$, and surface areas $|\partial D|$ and $|\partial E|$, then, for large c, the number $n(c,\alpha)$ of singular values of $P_{cD}\mathcal{F}P_{cE}$ greater than α is given approximately by

$$n(c,\alpha) \sim |D|\cdot|E|c^{2N} - \gamma|\partial D||\partial E|c^{2N-2}\log(\alpha^{-1}-1)\log c + o(c^{2N-3}) \qquad (2.10)$$

where cD is the set $\{cd: d \in D\}$ and γ is a constant independent of c, D and E.

ii. In one dimension let σ_n be the n-th singular value of $\mathcal{F}_c$. Then if b is fixed, c arbitrary and n is determined by

$$n = \left[4c^2 + \frac{2b}{\pi^2}\log(2c\sqrt{2\pi})\right] \qquad (2.11)$$

where $[\alpha]$ denotes the nearest integer to α, then

$$\lim_{n\to\infty} \sigma_n = (1+e^{b})^{-1/2} \quad . \qquad (2.12)$$

iii. The singular functions of $\mathcal{F}_c$ are the eigenfunctions of the Sturm-Liouville equation

$$((1-t^2)\phi')' + (\lambda - 4\pi^2 c^2 t^2)\phi = 0 \tag{2.13}$$

where solutions are required to be uniformly bounded over the entire real line.

The theorem indicates that the singular values σ_n of $P_D \mathcal{F} P_E$ have a steplike distribution; $\sigma_n \sim 1$ for small n, σ_n decays exponentially for large n, and the change from $\sigma_n \sim 1$ to exponential decay occurs over an interval of width proportional to $c^{2N-2} \log c$ centered on $c^2|D|\cdot|E|$. This in turn implies that the essential dimension is approximately $c^2|D|\cdot|E|$ and is almost independent of noise: The exponential decay of σ_n implies that the essential dimension $N = N(\delta, \varepsilon_1, \varepsilon_2)$ is determined by an equation of the form $e^{-N} \sim \varepsilon_1 \delta^{-1}$ so that the noise level ε_1 must be reduced by a multiplicative factor to give an additive increase in N.

A second consequence is that the low order singular functions are solutions to a Sturm-Liouville problem and therefore are analytic and slowly varying. Thus they will be well approximated by a discretization based on smooth functions. This was experimentally verified in [14], where a number of different discretizations of $\mathcal{F}_c$ were calculated for varying values of c. The results indicated that discretizations based on Gaussian quadrature or Galerkin approximations using Legendre polynomials required only $N + O(\log N)$ parameters to accurately approximate the first N singular functions of $\mathcal{F}_c$. In contrast Galerkin approximations based on piecewise constant or trigonometric functions appeared to require at least αN parameters to achieve the

same accuracy, where $\alpha \geq 3$. The difference may be explained by noting that the expansion of an analytic function on $[-c,c]$ in terms of Legendre polynomials will have rapidly decaying coefficients; whereas an expansion in terms of piecewise continuous functions will converge only slowly. Moreover although trigonometric functions are themselves smooth, they do not approximate the original function but rather a periodic extension of it outside of $[-c,c]$. This extension is likely to have discontinuities at the endpoints $\pm c$, so that its Fourier series is slowly convergent. This was observed in approximations of the singular functions of $\mathcal{F}_c$ where they had the worst performance of the discretizations examined; therefore their use is not recommended in transform recovery.

To conclude the section an example of solution of problem 1 by filtered SVD is presented, see [14] for the details and other examples. The equation

$$\mathcal{F}_c G = (P_c g)(v) + \varepsilon \mu(v) \tag{2.14}$$

was solved numerically, where

$$g(v) = 2\left(\frac{\sin \pi v}{\pi v}\right)^2 \quad . \tag{2.15}$$

This is the point spread function of an infocus, aberration free slit aperture. The presence of noise was simulated by the term $\varepsilon\mu(v)$ with $\mu(v)$ a random variable uniformly distributed over $[-1,1]$ and ε a control of the magnitude of the noise. The parameter values $c = 1$ and $\varepsilon = .03$ were chosen, a Galerkin approximation based on 80 piecewise constant functions was used to discretize the problem, and the resulting finite system was solved by filtered SVD using the filter of Eq. (2.9) with $q = 2$ and $\gamma = .01$.

Figure 1 shows two approximate solutions $\tilde{G}$ calculated from noise data along with the true solution for noiseless data

$$G(\omega) = 2(1 - |\omega|) \quad . \tag{2.16}$$

Figure 2 shows the extrapolation $\mathcal{F}\tilde{G}$ of one particular perturbation $P_c\tilde{g}$. Because of the symmetry of the test functions each graph is for negative values of the argument only.

The graphs show that $\tilde{G}$ is a good approximation to the true solution G, except at the origin where the smooth singular functions cannot reconstruct the discontinuity in slope. Since this discontinuity dominates the far field behavior of $\mathcal{F}G$, the extrapolation is not as accurate as the reconstruction.

3. SOURCES OF ILL-CONDITIONING IN PHASE RETRIEVAL

The previous discussion of ill-conditioning in the linear problem gives new insight into why the nonlinear problem of phase retrieval is ill-posed. Previous examinations of the problem have concentrated on showing that the problem is ill-posed due to violations of conditions 1 and 2 of Hadamard's. That it is also ill-posed due to violations of condition 3, and the implications such violations have for numerical solutions, has not been noted previous to [15]. We therefore present a brief summary of all three possible sources of ill-conditioning and their effects on the behavior of algorithms for numerical solution of such problems.

Violations of condition 1 are not in themselves important for the following reason. Phase retrieval is a model of a real world phenomenon known to exist. Therefore failure of the model to have a solution does not imply that the real physical quantity does not exist; so nonexistence must be due either to an inaccurate model or to noisy data. Both of these possibilities are simply extreme examples of discontinuous dependence of the solution on the data; therefore violations of condition 1 are subsumed under violations of condition 3.

Nonuniqueness, however, is an important source of ill-conditioning. The precise form of all possible solutions to the one-dimensional (1-D) phase problem appears to have been first determined by Akutowicz [27,28] and has been independently rediscovered by a number of other authors, e.g. [29,30]. Their results imply that the phase problem in 1-D has uncountably many

solutions. In [31], see also [15], we show that these results may be extended to two (and higher) dimensions to give necessary conditions on the form of multiple solutions for higher dimensions; these conditions imply that multiple solutions are significantly less likely than in 1-D.

The 1-D results depend on noting that any solution $g(v)$ is an analytic function of exponential growth: this follows from the fact that g is the transform of a function G with bounded support and the Paley-Wiener theorem. Therefore $g(v)$ has a Hadamard factorization [32] of the form

$$g(v) = |g(0)| e^{i(\alpha+\beta v)} \prod_{k=1}^{\infty} \left(1 - \frac{v}{v_k}\right) \tag{3.1}$$

where α and β are real constants and $\{v_k\}_{k=1}^{\infty}$ are the countably many zeroes of g. Then any other solution must be of the form

$$\tilde{g}(v) = e^{i(\tilde{\alpha}+\tilde{\beta}v)} B(v) g(v) \qquad \tilde{\alpha}, \tilde{\beta} \in \mathbb{R} \tag{3.2}$$

where $B(v)$ is a finite or infinite product of Blaschke factors, i.e.

$$B(v) = \prod_{\ell=1}^{\infty} B_{k_\ell}(v) \quad \text{where} \quad B_k(v) = \frac{v-v_k^*}{v-v_k} \quad . \tag{3.3}$$

Furthermore if $\tilde{\beta} = 0$ then any $\tilde{g}$ given by Eq. (3.2) is indeed a solution. Thus alternative exact solutions are basically generated by "flipping" zeroes of $g(v)$ to their complex conjugates. Since there are an infinite number of zeroes there are also an infinite number of exact solutions to a 1-D phase retrieval problem.

If the exact solutions were few and well separated, and the phase retrieval problem well-posed in a neighborhood of each zero, then any standard numerical algorithm would perform satisfactorily. However, as noted by Napier [33], any N zeroes may be flipped or not flipped in 2^N different combinations, giving 2^N different solutions. Therefore there is a very large number of possible solutions, in fact an uncountable infinity of such solutions. Furthermore, for reasonable functions $G(\omega)$, any infinite product $B(z)$ of Blaschke factors will converge, i.e. $B(z) = \lim_{N\to\infty} B_N(z)$ where $B_N(z)$ is a product of N Blaschke factors. Since each finite product corresponds to a possible solution, it follows that the set of solutions has limit points. *Any numerical algorithm will have great difficulty in the neighborhood of such points.*

However the situation improves markedly in $N \geq 2$ dimensions. An extension of the arguments for $N = 1$ shows that the zeroes of any alternative solution $\tilde{g}$ are the zeroes, or complex conjugates of the zeroes, of g. But, as g is an entire function of N complex variables, its zeroes form an analytic set X of dimension $N-1$ [34]. This set X is essentially the union of M connected, N-1 dimensional, analytic manifolds X_i. Therefore, if part of a manifold X_i is flipped to form the zeroes of $\tilde{g}$, all of X_i must be flipped to ensure that the zeroes of $\tilde{g}$ also form an analytic set $\tilde{X}$; so $\tilde{X}$ will be of the form $\bigcup_{k=1}^{P} X_{i_k} \cup \bigcup_{\ell=1}^{M-P} X^*_{i_\ell}$. But in two or more dimensions the set X is likely to be irreducible, i.e. $M = 1$, in the same way that almost all polynomials of two or more variables are irreducible. Therefore at most two possible solutions with zeroes X and X^* can be

formed, and the solution is essentially unique.

Therefore, in the simplest model problem, ill-conditioning due to non-uniqueness is likely to be a severe problem in one-dimensional problems, or in problems that are essentially one-dimensional (e.g. those with radial symmetry considered in [35]); but in higher dimensions it should have significantly less effect. The situation is less well understood if side conditions are imposed. For instance if both g and G are analytic then the solution of example 2a is essentially unique, but for arbitrary G, nonunique solutions have been constructed (see [11] for a review). Even less is known about the effect of positivity.

However, it should be noted that for all three problems there can be parasitic solutions due to symmetries, etc. For example, if G vanishes outside of the interval dI where $d < c$ then translations $\tilde{G}$ of G will still be within cI and will have transforms $\tilde{g}$ which have the same modulus as g over cI.

Finally we show that example 2 is locally ill-conditioned in that it violates condition 3. The phase retrieval problem may be recast as requiring the solution of the nonlinear equation

$$\mathscr{L}(G) = m \qquad \mathscr{L}: L^2(B) \to L^2(A) \tag{3.4}$$

where $\mathscr{L}$ is the composite operator $\mathscr{L}_1 \circ P_A \mathscr{F} P_B$ and

$$\mathscr{L}_1(g) = |g| \qquad \mathscr{L}_1: L^2(A) \to L^2(A) \quad . \tag{3.5}$$

Since $\mathscr{L}_1$ is a bounded continuous operator and $P_A \mathscr{F} P_B$ is compact, $\mathscr{L}$ is

compact. This implies that for any $\varepsilon > 0$ an infinite sequence of functions $\{G_i\}$ can be found, such that

$$\|G_i - G_j\| > 1 - \delta_i ; \qquad \text{but} \quad \|\mathscr{L}(G_i) - \mathscr{L}(G_j)\| < \varepsilon \quad . \tag{3.6}$$

Nonuniqueness led to global ill-conditioning, in that there are regions in which many exact solutions exist. Compactness leads to local ill-conditioning, in that in the neighborhood of an exact solution there are directions H in which a change in the solution G induces a negligibly small change in the observation m. Therefore although $\mathscr{L}(G)$ and $\mathscr{L}(G+H)$ are distinct in theory in practice, with the presence of measurement noise, they are indistinguishable.

This local ill-conditioning may be partially quantified by noting that compactness is due to the operator $P_A \mathscr{F} P_B$, and that inversion of $\mathscr{L}$ involves inversion of $P_A \mathscr{F} P_B$. Therefore the theory of the essential dimension and the idea of filtered inversion outlined in the previous section suggest that the essential dimension of the phase problem (the number of complex parameters that may be accurately determined) is approximately bounded above by $N = (a_2 - a_1) \cdot (b_1 - b_2)$, is relatively independent of the noise level in m, and that the solution space should be restricted by filtering to the span of the first N singular functions of $P_A \mathscr{F} P_B$. Furthermore the discretization chosen for the problem should accurately approximate these singular values.

The local ill-conditioning of examples 2a and 2b is less well understood. However it is reasonable to suppose that the total amount of information available in all the constraints is less than the maximum amount of

information in each constraint considered separately. Therefore, if G is represented by $2P$ real parameters, in example 2a P of these are determined by knowledge of $|G|$, and up to $2N$ by knowledge of $|g|$, giving an upper bound on the essential dimension of $P + 2N$. In example 2b the upper bound is $P + N$; in this case the condition that G be real places symmetry constraints on g, effectively halving the amount of information available in m. Again these results suggest that numerical solutions be constructed in a similar fashion as solutions to the linear problem.

4. ITERATED PROJECTION ALGORITHMS

The previous sections showed that transform recovery problems are locally ill-conditioned, which confirms the practical experience of a number of authors [1,8,36] who noted very slow convergence rates of the Gerschberg-Saxton (GS) algorithm. Therefore, in order to modify the algorithm to cope with this ill-conditioning, we place GS in a more general setting by viewing it as a special case of finding a common intersection point of a collection of sets. In formal language, it is:

Given the sets $\{T_i\}_{i=1}^{M}$ with associated projections $P_i \equiv P_{T_i}$

find G such that $G \in \bigcap_{i=1}^{M} T_i$

Gubin, Polyak and Raik [37] have proposed an iterative algorithm for the solution of such problems in which at the n-th stage G_{n+1} is generated by

$$G_{n+1} = P_i G_n \qquad \text{where} \quad i = (n-1) \bmod M + 1 \tag{4.1}$$

and proved that under certain conditions, such as the convexity of the sets T_i, the iterates converged to a common intersection point if one existed. A survey of this and similar algorithms appearing in the Russian literature is given by Censor and Herman in [38]. If the set S is defined to be

$$S = \{g \in L^2(\mathbb{R}): \; g(v) = \tilde{g}(v) \quad \text{for} \quad v \in A\} \tag{4.2}$$

for example 1, or

$$S = \{g \in L^2(\mathbb{R}): \; |g(v)| = m(v) \quad \text{for} \quad v \in A\} \tag{4.3}$$

for example 2, and the sets T_1 and T_M defined as

$$T_1 = \{G \in L^2(\mathbb{R}): \quad G = \mathcal{F}^{-1} g, \; g \in S\} \tag{4.4}$$

$$T_M = \{G \in L^2(\mathbb{R}): \quad G(\omega) = 0 \quad \text{for} \quad \omega \notin B\} \tag{4.5}$$

then, as the Fourier transform preserves L^2 norms

$$h = P_S g \Longleftrightarrow H = \mathcal{F}^{-1} h = P_1(\mathcal{F}^{-1} g) = P_1 G \tag{4.6}$$

and GS applied to examples 1 and 2 is recognizable as the iterated projection algorithm of Eq. (4.1) with $M = 2$.

The advantages of the iterated projection algorithm are: First that it allows easy incorporation of extra constraints such as those in examples 2a and 2b by setting $M = 3$ and adding either the set

$$T_2 = \{G \in L^2(\mathbb{R}): \; G(\omega) \geq 0 \quad \text{for} \quad \omega \in B\} \tag{4.7}$$

or

$$T_2 = \{G \in L^2(\mathbb{R}): \; |G(\omega)| = n(\omega) \quad \text{for} \quad \omega \in B\} \; . \tag{4.8}$$

Second that in transform recovery problems the projections P_i may be very easily computed. The disadvantage is that, as stated, the algorithm is sensitive to local ill-conditioning. Figure 3 shows two instances of the effects of ill-conditioning: in the first the sets intersect at a very acute angles so that the projections are very slowly convergent, and in the second the presence of noise has perturbed the two sets so that an intersection point does not exist. These possibilities, and the additional fact that T_1 is not convex in phase retrieval problems, imply that the iterated projection algorithm will either be very slowly convergent or fail to converge at all.

In order to escape these difficulties we propose that the original problem be replaced by

$$\text{Find } G \text{ to minimize } F(G) = \sum_{i=1}^{M} \|G - P_i G\|^2 \quad .$$

Obviously $F(G) \geq 0$ and $F(G) = 0$ iff G is a common intersection point, but even if such a point does not exist due to perturbations of the sets by noise in the data, the $\tilde{G}$ that minimizes $F(G)$ is an acceptable pseudo-solution to the problem. We also propose the following extension of the iterative projection algorithm for minimization of $F(G)$ in which at the n-th iteration

$$\begin{aligned} H_{n+1} &= \frac{1}{(M-1)} \sum_{i=1}^{M-1} (P_i G_n - G_n) \\ G_{n+1} &= P_M(G_n + \lambda_n H_n) \end{aligned} \tag{4.9}$$

where

$$\begin{aligned} \lambda_M &\in (0,2) \quad \text{if } T_M \text{ convex} \\ &\in (0,1) \quad \text{otherwise} \end{aligned} \quad . \tag{4.10}$$

This algorithm will be termed the *restricted projection* (RP) *algorithm* from hereon.

RP has a number of useful features, in particular it produces constantly decreasing residuals as does GS [1], which are summarized in the following Theorem, proved in [15].

Theorem 4.1. If the projections $P_i G$ are unique and continuous in G then at the n-th iteration the iterates G_n of Eq. (4.9) satisfy

$$F(G_{n+1}) < F(G_n) \quad \text{or} \quad G_{n+k} = G_n \qquad \forall k \geq 1 \quad . \tag{4.11}$$

Furthermore if $\tilde{G}$ is a limit point of the iterates then $\tilde{G}$ is a fixed point of the iteration.

RP can also be recast as one of the standard optimization algorithms as the next theorem from [15] shows.

Theorem 4.2. The gradient $\nabla F(G)$ of $F(G)$ is

$$\nabla F(G) = 2 \sum_{i=1}^{M} (G - P_i G) \quad . \tag{4.12}$$

Therefore if $T_M = L^2(\mathbb{R})$ then RP is the standard steepest descent algorithm with variable steplength $\lambda \in (0, \frac{1}{M-1})$.

We have presented the algorithm in a form in which iterates are restricted to the particular set T_M. This includes the unrestricted case ($T_M = L^2(\mathbb{R})$), but also allows from knowledge about the solution G to be reimposed at each iteration after less well-known requirements are satisfied by moving in the search direction. However, the chief benefit of the restriction in transform recovery problems with the set T_M of Eq. (4.5) is that the resulting algorithms are *efficient*. That is at each iteration they require function values of G and $\mathcal{F}G$ only across the intervals A and B. Efficiency has not always been achieved, for instance the version of GS proposed by Papoulis [2] for inversion of the fFT requires knowledge of G_n across the entire line at each iteration.

To demonstrate that transform recovery problems are efficient we first note that if T_M is a linear subspace then the projection P_M is linear and

$$P_M(G_n + \lambda H_n) = G_n + \frac{\lambda}{M-1} \sum_{i=1}^{M-1} P_M P_i G_n \tag{4.13}$$

so that the search direction is independent for λ, and for any λ the new iterate is still in T_M. Now for convenience let $A = B = cI$ so that $P_M = P_c$, then

$$P_c P_1 G = P_c \mathcal{F}^{-1} F_S \mathcal{F} G \quad . \tag{4.14}$$

$P_S g$ is easily shown to be

$$P_S g = \begin{cases} m(v) e^{i \arg g(v)}, & v \in cI \\ g(v), & \text{otherwise} \end{cases} \tag{4.15}$$

so that

$$P_S y = g - P_c g + P_c P_S P_c g \tag{4.16}$$

and

$$P_c P_1 G = G - \mathcal{F}_c^* \mathcal{F}_c G + \mathcal{F}_c^* P_S \mathcal{F}_c G \quad . \tag{4.17}$$

Moreover for example 2a

$$(P_c P_2 G)(\omega) = \begin{cases} n(\omega) e^{i \arg G(\omega)}, & \omega \in cI \\ 0, & \text{otherwise} \end{cases} \tag{4.18}$$

and for example 2b

$$(P_c P_2 G)(\omega) = \begin{cases} (\text{Re } G)(\omega), & \text{if } (\text{Re } G)(\omega) > 0 \text{ and } \omega \in cI \\ 0 & \text{otherwise} \end{cases} \tag{4.19}$$

It is obvious that calculation of $P_c P_i G$ requires only values of G and $g = \mathcal{F} G$ on cI, so that RP is efficient.

We conclude this section by noting that the relationship between RP and gradient method opens up new possibilities for improvement of RP. In particular, as noted in Eq. (4.13), if T_M is linear then the search direction is independent of λ, therefore it is possible to do a line search in the direction of H_n. By Eq. (4.12)

$$\frac{d}{d\lambda} F(G_n + \lambda H_n) = \nabla F(G_n + \lambda H_n, H_n)$$

$$= 2 \sum_{i=1}^{M-1} (G_n + \lambda H_n - P_i(G_n + \lambda H_n), H_n) \quad . \tag{4.20}$$

Since calculation of $F(G)$ requires values of $P_i G$, calculation of $F(G_n + \lambda H_n)$ at any point in a line search gives sufficient information for calculation of the gradient at that point. Therefore a line search algorithm using derivative information, such as Powell's cubic line search algorithm [39], may be used for the same cost as a standard quadratic line search that uses values of $F(G)$ only. Since in phase retrieval $T_M = L^2(cI)$, cubic line searches may be profitably employed.

5. ALGORITHMS BASED ON AFFINE APPROXIMATIONS

The previous section showed that the GS algorithm could be extended to an algorithm RP that mitigated some of the effects of ill-conditioning in transform recovery. However the extension does not remove all of these effects as, in particular, it performs no filtering to restrict the solution to a well-posed solution set. Moreover as RP is a variant of steepest descent, it is only of first order and therefore will not perform well even on some well-posed problems for the same reasons that gradient algorithms perform poorly on some standard optimization problems. Therefore we seek a solution to both these problems by proposing a new class of iterative algorithms based on more accurate affine approximations to the sets T_i or the functions $F(G)$ and P_iG. At each iteration these algorithms require the solution of an ill-posed linear subproblem similar to that discussed in Section 2; this may be done using filtered SVD thus further reducing ill-conditioning in phase retrieval.

We start with a simple example of how such algorithms may be constructed. In RP the search direction H_n may be viewed as being obtained by replacing the sets T_i by point approximations P_iG_n at the n-th iteration, and then solving the subproblem

$$\text{Minimize } F_n(G) = \sum_{i=1}^{M} \|G - P_iG_n\|^2 \quad . \tag{5.1}$$

In particular the set S is approximated by the point P_Sg_n. However the only restrictions on functions in S is that they have value $\tilde{g}$ or modulus

m over the interval A; outside of A they may take on arbitrary values. Therefore, if $A = B = cI$, the affine subspace

$$S_n = \{h: h = P_c P_S g_n + (g - P_c g), g \in S\} \tag{5.2}$$

contains the point $P_S g_n$ and is contained in the set S. If it is used to replace the point approximation $P_1 G_n$ to T_1 by $T_{1n} = \mathcal{F}^{-1} S_n$ at the n-th iteration, then the new subproblem to be solved is

$$\text{minimize } F_n(G) = \|G - P_{T_{1n}} G\|^2 + \sum_{i=2}^{M-1} \|G - P_i G_n\|^2 \quad . \tag{5.3}$$

The minimum L_n satisfies the normal equations

$$(\mathcal{F}_c^* \mathcal{F}_c + (M-2)\mathcal{I}) L = \mathcal{F}_c^* P_S g_n + \sum_{i=2}^{M-1} P_i G_n \tag{5.4}$$

giving a search direction $H_n = L_n - G_n$. If $M > 2$ then this subproblem is well-posed as $(\mathcal{F}_c^* \mathcal{F}_c + (M-1)\mathcal{I})$ has a bounded inverse; but if $M = 2$ then the problem is inversion of the fFT $\mathcal{F}_c$ recast as a linear least squares problem. This is an ill-posed problem best solved by filtered SVD as described in Section 2; but as the same linear operator appears at each iteration the SVD need be calculated only once.

A more accurate approximation arises from replacing the projections $P_i G$ by the linear approximation

$$P_i G \sim P_i G_n + \mathcal{H}_i(G_n)(G - G_n) \tag{5.5}$$

at the n-th iteration, where $\mathcal{H}_i$ is the Frechet derivative of the operator P_i. We assume that $\mathcal{H}_i$ exists and is a bounded linear operator; conditions

on the sets T_i that would guarantee these properties are quite complicated (e.g. [40]) and are beyond the scope of this work. However if $\mathcal{H}_i$ possesses these properties then it is symmetric, $\mathcal{I}-\mathcal{H}_i$ is the Hessian (i.e. the second Frechet derivative) of the function $F_i(G) \equiv \|G - P_i G\|^2$ and for every G

$$(\mathcal{I}-\mathcal{H}_i(G))(G - P_i G) = 0 \quad . \tag{5.6}$$

If this approximation is used at the n-th iteration the corresponding subproblem to be solved is

$$\text{minimize } F_n(G) = \sum_{i=1}^{M-1} \|G - P_i G_n - \mathcal{H}_i(G_n)(G - G_n)\|^2 \quad . \tag{5.7}$$

The function L that solves this problem is by definition the least squares solution to the block system

$$\begin{bmatrix} \mathcal{I}-\mathcal{H}_1(G_n) \\ \vdots \\ \mathcal{I}-\mathcal{H}_M(G_n) \end{bmatrix} L = -\begin{bmatrix} G_n - P_1 G_n \\ \vdots \\ G_n - P_{M-1} G_n \end{bmatrix} \tag{5.8}$$

Since the original transform recovery problem was ill-posed, this block system is also usually ill-posed and should be inverted by filtered SVD. However the cost of calculation of the SVD of the entire block matrix at each iteration would very expensive. Therefore we propose that the SVD of each block $\mathcal{I}-\mathcal{H}_i(G_n)$ be calculated and filtered separately to give an approximate filtering of the whole matrix. As we shall see later in the section such filtering can be done relatively cheaply.

The third set of algorithms described here are based on the following affine approximation of $F(G)$ by

$$F(G) \sim F(G_n) + \nabla F(G_n)(G - G_n) + \frac{1}{2}(G - G_n)\nabla^2 F(G_n)(G - G_n) \quad . \tag{5.9}$$

From Eqs. (4.11) and (5.5)·

$$\nabla F(G) = 2 \sum_{i=1}^{M} (G - P_i G)$$

$$\nabla^2 F(G) = 2 \sum_{i=1}^{M} (\mathcal{I} - \mathcal{H}_i(G)) \quad . \tag{5.10}$$

At each iteration this approximation gives the subproblem

$$\text{minimize } F_n(G) = F(G_n) + \nabla F(G_n)(G - G_n) + \frac{1}{2}(G - G_n)\nabla^2 F(G_n)(G - G_n) \tag{5.11}$$

which, if $\nabla^2 F(G_n)$ is positive definite, has as its unique minimum the solution L_n of Newton's equation

$$\nabla^2 F(G_n) L = -\nabla F(G_n) \tag{5.12}$$

giving the standard Newton search direction $H_n = L_n - G_n$. However, for phase retrieval problems, in order that the resulting algorithm be efficient in the sense of the previous section we replace Eq. (5.12) by the restricted equation

$$P_c \nabla^2 F(G_n) P_c L = -P_c \nabla F(G_n) \tag{5.13}$$

so that values of iterates are required only over the interval cI.

Equations (5.12) and (5.13) are normally ill-posed and therefore best solved by filtered SVD. As the fFT, with its exponentially decaying singular values, underlies the Hessian $\nabla^2 F(G)$ we propose a cutoff filter of the form

$$\begin{aligned} f(\lambda) &= \lambda^{-1} && \text{if } |\lambda| \geq \varepsilon_3 \\ &= 0 && \text{otherwise} \end{aligned} \tag{5.14}$$

where ε_3 is dependent on the noise levels and desired accuracy. This should produce a search direction H_n in which $F(G)$ should vary moderately rapidly as directions corresponding to small eigenvalues, and thus slowly varying F, have been filtered out. However the resulting direction is not necessarily a descent direction, as for some G, $\nabla^2 F(G)$ will have negative eigenvalues; but this may be corrected by use of the filter

$$\begin{aligned} f(\lambda) &= \lambda^{-1} && \text{if } \lambda \geq \varepsilon_3 \\ &= 0 && \text{otherwise} \end{aligned} \quad . \tag{5.15}$$

Newton's algorithm in which Eq. (5.13) is inverted with the filter of Eq. (5.14) shall be denoted by FN; if the filter of Eq. (5.15) is used it will be denoted by FNP.

Since $\nabla^2 F(G_n)$ changes with each iteration FN and FNP incur considerable costs in calculation of a new SVD at each iteration. Although some savings are possible by using the old SVD as a first approximation to the new SVD with optional iterative refinement, we instead sought to reduce Eq. (5.13) to the block form of Eq. (5.8) so that block filtering could be applied. This may be done by noting that if $\mathscr{I}-\mathscr{K}_i$ is a symmetric positive

definite square root $(\mathcal{I} - \mathcal{H}_i)^{1/2}$ which by Eq. (5.6) satisfies

$$(\mathcal{I} - \mathcal{H}_i(G))(G - P_i G) = G - P_i G \quad . \tag{5.16}$$

Therefore Eq. (5.13)

$$\left[\sum_{i=1}^{M-1} P_c(\mathcal{I} - \mathcal{H}_i(G_n))P_c\right]L = -\sum_{i=1}^{M-1}(G_n - P_c P_i G_n) \tag{5.17}$$

may be rewritten as

$$\left[\sum_{i=1} P_c((\mathcal{I} - \mathcal{H}_i(G_n))^{1/2})^2 P_c\right]L = -\sum_{i=1}^{M-1} P_c(\mathcal{I} - \mathcal{H}_i(G_n))^{1/2}(G_n - P_i G_n) \tag{5.18}$$

which in turn is recognizable as the normal equations for the least squares solution of the block system

$$\begin{bmatrix} (\mathcal{I} - \mathcal{H}_1(G_n))^{1/2} P_c \\ \vdots \\ (\mathcal{I} - \mathcal{H}_{M-1}(G_n))^{1/2} P_c \end{bmatrix} L = -\begin{bmatrix} G_n - P_c P_1 G_n \\ \vdots \\ G_n - P_c P_{M-1} G_n \end{bmatrix} \quad . \tag{5.19}$$

Block filtering was chosen because of the simple form of the blocks in Eq. (5.19) in phase retrieval problems. If the operators P_i are viewed as acting on the pair of real functions $\begin{bmatrix} \text{Re } H(\omega) \\ \text{Im } H(\omega) \end{bmatrix}$ instead of on the complex valued function $H(\omega)$ then simple calculations show that $\mathcal{I} - \mathcal{H}_2(G)$ may be represented as a block diagonal operator whose 2×2 diagonal blocks, are indexed by the variable ω. For example 2a

$$(\mathcal{I} - \mathcal{H}_2(G))(\omega) = \begin{bmatrix} \sin\theta & \cos\theta \\ -\cos\theta & \sin\theta \end{bmatrix} \begin{bmatrix} 1 - \frac{n(\omega)}{|G(\omega)|} & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \sin\theta & -\cos\theta \\ \cos\theta & \sin\theta \end{bmatrix} \tag{5.20}$$

where $\theta = \arg G(\omega)$

and for example 2b

$$(\mathcal{I} - \mathcal{H}_2(G))(\omega) = \begin{cases} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} & \text{if } (\mathrm{Re}\ G)(\omega) > 0 \text{ and } \omega \in cI \\ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} & \text{otherwise} \end{cases} \tag{5.21}$$

The spectrum of each operator $\mathcal{I} - \mathcal{H}_2$ is now recognizable as the union over ω of the spectrum of each block. For example 2b the block spectra consists only of the set $\{0,1\}$ so $\mathcal{I} - \mathcal{H}_2$ is its own square root. For example 2a the spectrum of $\mathcal{I} - \mathcal{H}_2$ takes on a range of values, some of which may be small or negative. Consequently $(\mathcal{I} - \mathcal{H}_2)^{1/2}$ is either ill-conditioned, or not well defined (i.e. has imaginary eigenvalues) or is both. Therefore a filter with one of the following forms is used

$$\begin{aligned} f(\lambda) &= |\lambda|^{1/2} \operatorname{sign} \lambda, && \text{if } |\lambda| \geq \varepsilon_3 \\ &= 0, && \text{otherwise} \end{aligned} \tag{5.22}$$

$$\begin{aligned} f(\lambda) &= \lambda^{1/2}, && \text{if } \lambda \geq \varepsilon_3 \\ &= 0, && \text{otherwise} \end{aligned} \tag{5.23}$$

to replace the entry $1 - \frac{n(\omega)}{|G(\omega)|}$ in Eq. (5.20) by $f(1 - \frac{n(\omega)}{|G(\omega)|})$. Finally for either problem the original block $(\mathcal{I} - \mathcal{H}_2(G))P_c$ is reduced in size by

eliminating from the block those equations for $(\text{Re } H)(\omega)$ or $\sin\theta(\text{Re } H)(\omega) - \cos\theta(\text{Im } H)(\omega)$ that correspond to eigenvalues that have been filtered to zero.

From Eq. (4.15)

$$P_c(\mathcal{I} - \mathcal{H}_1(G))P_c = \mathcal{F}_c^*(\mathcal{I} - \mathcal{H}_S(g))\,\mathcal{F}_c \tag{5.24}$$

where $g = \mathcal{F}G$ and $\mathcal{I} - \mathcal{H}_S(g)$ is the block diagonal operator

$$(\mathcal{I} - \mathcal{H}_S(g)) = \begin{bmatrix} \sin\phi & \cos\phi \\ -\cos\phi & \sin\phi \end{bmatrix} \begin{bmatrix} 1 - \frac{m(v)}{|g(v)|} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \sin\phi & -\cos\phi \\ \cos\phi & \sin\phi \end{bmatrix} \tag{5.25}$$

where $\phi = \arg g(v)$.

$P_c(\mathcal{I} - \mathcal{H}_1)P_c$ does not possess an obvious symmetric square root however Eq. (5.24) suggests the asymmetric square root $(\mathcal{I} - \mathcal{H}_S(g))^{1/2}\mathcal{F}_c$. This composite operator may be approximately filtered by filtering each component; $(\mathcal{I} - \mathcal{H}_S(g))^{1/2}$ may be filtered as was $(\mathcal{I} - \mathcal{H}_2)^{1/2}$ in example 2a and $\mathcal{F}_c$ may be filtered after calculation of its SVD as described in Section 2. Since the first filtering requires a trivial calculation at each iteration, and $\mathcal{F}_c$ is independent of the iterates, this filtering may again be done cheaply.

Therefore, after rows corresponding to filtered elements have been eliminated, in phase retrieval problems we are left with a reduced block system

$$\begin{bmatrix} (\mathcal{I} - \mathcal{H}_S(g))_f^{1/2}(\mathcal{F}_c)_f \\ (\mathcal{I} - \mathcal{H}_2(G)_f^{1/2} \end{bmatrix} L = \begin{bmatrix} P_c P_S \mathcal{F}_c G_n \\ P_2 G_n - G_n \end{bmatrix}_f \tag{5.26}$$

to solve in a least squares sense. This can be done using a standard routine such as LLSQF in the IMSL library, or further advantage may be taken of sparsity within the system. If Eq. (5.26) is solved at each iteration with the filter of Eq. (5.22) resulting algorithm is termed SQFN; if the filter of Eq. (5.23) is used the algorithm is SQFNP.

6. SOME ILLUSTRATIVE NUMERICAL RESULTS

To conclude the paper we give a brief account of some numerical solutions to examples 2, 2a and 2b using the algorithms proposed in the previous sections. Reference is made to our two previous papers for detailed numerical computations of a wide selection of problems. The measure of performance of the algorithms was taken to be the number of iterations required to reduce the function $F(G_n)$ to below a prescribed value. The ill-posed nature of phase retrieval problems implies that this is not the best of measures in that a small value $F(G_n)$ does not necessarily imply that G_n is close to the true solution G. However without knowledge of the true solution we have none better. Therefore, because of this ill-conditioning and the limited resources for numerical computation at our disposal, the results presented here are intended as a guide to the behavior of the algorithms on real problems rather than a firm prediction of their likely performance.

We begin with an account of the mechanics of the computation. First is the discretization: it was determined by the requirements that it gives an accurate, efficient approximation to $\mathcal{F}_c$, and that the discrete projections be easily calculated. Therefore a discretization based on N point Gaussian quadrature was chosen; the numerical experiments mentioned in Section 2 showed its efficiency and its pointwise nature allows easy evaluation of projections. The discrete problem thus involves determination of vectors $\hat{g}$, $\hat{G} \in C^N$ whose elements g_k, G_k are to be approximations to the function values $g(\rho_k)$, $G(\rho_k)$ at the abscissae ρ_k of the N-point Gaussian quadrature

rule on cI. To this end vectors $\hat{m}$, $\hat{n} \in \mathbb{R}^N$ are formed with components $m_k = m(\rho_k)$, $n_\ell = n(\rho_\ell)$ where $m(v)$ and $n(\omega)$ are the known moduli. Then matrices $\hat{W}$, $\hat{F} \in C^{N\times N}$ are constructed with $\hat{W}$ being a diagonal matrix whose k-th entry is the weight w_k of the quadrature rule, and $\hat{F}$ having entries $F_{k\ell} = e^{2\pi i \rho_k \rho_\ell}$. The discretized example problems are now.

Find vectors $\hat{g}$, $\hat{G}$ such that $\hat{g} = \hat{F}\hat{W}\hat{G}$ and

Example 2: $|g_k| = m_k$

Example 2a: $|g_k| = m_k$, $|G_\ell| = n_\ell$

Example 2b: $|g_k| = m_k$, $G_\ell \geq 0$.

In order to test the algorithms, particular examples of each model problem were chosen. Of particular interest, in-as-far as this paper is concerned, is the test function

$$G(\omega) = e^{i2\pi W(\frac{\omega}{c})} \tag{6.1}$$

$$W(\tfrac{\omega}{c}) = W_3 S_3(\tfrac{\omega}{c}) + W_4 S_4(\tfrac{\omega}{c})$$

$$= W_3[(\tfrac{\omega}{c})^3 - \tfrac{3}{5}(\tfrac{\omega}{c})] + W_4[(\tfrac{\omega}{c})^4 - \tfrac{6}{7}(\tfrac{\omega}{c})^2] \tag{6.2}$$

$G(\omega)$ is the pupil function of a slit aperture having unit amplitude over the exit pupil and suffering from wavefront aberrations of optimum balanced coma S_3 and optimum balanced spherical aberration S_4, where W_3/λ and W_4/λ are the dimensionless aberration strengths. S_3 and S_4 are the slit aperture versions [41] of the Zernike polynomials. This test function was used previously in [9]. In reference [9], the inversion assumed that the amplitude distribution over the exit pupil was unity; in the present case

neither the amplitude distribution or wavefront are known *a priori*. The numerical calculations summarized here were carried out for $W_3 = W_4 = 3\lambda/8$.

We now outline the basic structure of all the numerical test runs before considering various points in detail. All tests consisted of the four essential steps:

1. Choose an initial guess $\hat{G}_0$.
2. Given iterates $\hat{g}_n$ and $\hat{G}_n$ compute a search direction $\hat{H}_n$.
3. Calculate a steplength λ_n and new iterates

$$\hat{G}_{n+1} = \hat{G}_n + \lambda_n \hat{H}_n, \qquad \hat{g}_{n+1} = \hat{F}\hat{W}\hat{G}_{n+1} .$$

4. Iterate steps 2 and 3 until the convergence critiria are satisfied.

The convergence criteria used in all calculations were

$$F(\hat{G}_n) < 10^{-5} \qquad \text{or} \qquad \sum_{k=0}^{2} \|\lambda_{n-k}\hat{H}_{n-k}\| < 2 \times 10^{-3} \tag{6.3}$$

together with an upper limit N_{max} on the number of iterations. The sum of the last three steplengths, rather than $\|\lambda_n \hat{H}_n\|$ alone, was chosen as in ill-conditioned minimization problems "stop-start" behavior is often noticed. That is a large step often followed by one or two small steps, after which another large step is taken. This feature was often observed in the use of affine approximation algorithms, differences in magnitude of successive steplengths by factors greater than 100 occurred fairly frequently.

Three different forms of initial guess were used:

1. $\hat{G}_0 = 0$,
2. $\hat{G}_{0\ell} = (1-\varepsilon_1)(1+\varepsilon_1 r_\ell) G_\ell^i$,

3. $\hat{G}_{0\ell} = 10^{-3} r_\ell$.

Guess 2 is a damped perturbation of the true solution $\hat{G}^i$ with ε_1 representing the noise level. For convenience we shall express this level as a percentage, e.g. $\varepsilon_1 = .2$ will be described as $\varepsilon_1 = 20\%$ noise. The variable r_ℓ is a random complex variable with modulus uniformly distributed over $[0,1]$ and phase uniformly distributed over $[0,2\pi]$. Guess 3 represents a small totally random perturbation about the origin; again for convenience such guesses shall be denoted by $\varepsilon_1 = 100\%$.

The first results reported are those on determination of an optimal choice of λ_n. Three possibilities were considered:

1. $\lambda_n^1 = 1$.
2. $\lambda_n^2 = r_n$ where r_n is a random variable uniformly distributed over $[0,2]$.
3. λ_n^3 is the approximate minimum of $F(\hat{G}_n + \lambda_n \hat{H}_n)$ as a function of λ determined by Powell's cubic line search algorithm [39] with the following convergence criteria on the iterates ${}_k\lambda_n^3$:

$$\left|\frac{\left({}_k\lambda_n^3 - {}_{k-1}\lambda_n^3\right)}{{}_k\lambda_n^3}\right| \cdot \left|\frac{F\left(\hat{G}_n + {}_k\lambda_n^3 \hat{H}_n\right)}{F(\hat{G}_n)}\right| < .02$$

$$\left|\frac{\nabla F\left(\hat{G}_n + {}_k\lambda_n^3 \hat{H}_n\right)}{\nabla F(\hat{G}_n)}\right| \cdot \left|\frac{F\left(\hat{G}_n + {}_k\lambda_n^3 \hat{H}_n\right)}{F(\hat{G}_n)}\right| < .02 \qquad k \leq 5 \quad . \tag{6.4}$$

The performances of λ_n^i were compared by running each possibility on each model problem with the appropriate test functions using RP. The parameters

$c = 2$, $N = 40$ and $N_{max} = 50$ were chosen and each problem was started with three different initial guesses with $\varepsilon_1 = 20\%$, 60% and 100%.

The results were remarkably uniform over all test cases of model problem, test function and initial guess. Choices λ_n^1 and λ_n^2 performed almost identically with λ_n^3 just under a factor of 2 better. Almost always only one extra function evaluation was required for λ_n^3. This extra computation almost exactly balances the savings in the reduced number of iterations so that all three choices incurred the same computational cost in reduction of $F(\hat{G}_n)$ to a specified value. However the greater flexibility of λ_n^3 led to its adoption in all subsequent calculations.

We next attempted to estimate the local ill-conditioning in the problem by adding a small perturbation of size ε_2 to the data and starting the algorithm at the true solution $\hat{G}^i$ of the unperturbed problem. The results were inconclusive; although the algorithms terminated at a vector $\tilde{\hat{G}}^i$ such that $\|\tilde{\hat{G}}^i - \hat{G}^i\|/\|\hat{G}^i\| \sim \varepsilon_2$, this was due to the conditions $F(\tilde{\hat{G}}^i) < 10^{-5}$ or $n > N_{max}$ being satisfied, and not due to convergence of the iterates which appeared to cycle around some fixed point.

Further results showed that the problem was globally ill-conditioned in that a number of differing functions $\tilde{\hat{G}}^i$ were found starting from a random guess ($\varepsilon_1 = 100\%$) such that $F(\tilde{\hat{G}}^i) \sim 10^{-4}$ for example 2, or $F(\tilde{\hat{G}}^i) \sim 10^{-2}$ for examples 2a and 2b, but with $\|\tilde{\hat{G}}^i - \hat{G}^i\|/\|\hat{G}^i\| \sim 1$. This suggests that the surface $\{(F(G),G): G \in cI\}$ is very rugged, as expected from the results in Section 3 on existence of multiple solution in one dimension. The results also showed that the iterates determined as approximate solutions to example

2a and 2b were definitely more acceptable as solutions than those for example 2, even though as measured by $F(G)$ they were worse by a factor of 10 or more.

A good initial guess is of great help. If one is not available then it is tempting to start with guess i, however some analysis [9] has shown that if the iterates have some symmetries, then such symmetries will be preserved by the algorithm regardless of the form of the true solution. Thus the highly symmetric choice of $\hat{G}_0 = \hat{0}$ is to be avoided.

An attempt to estimate the effects of ill-conditioning due to the presence of $\mathscr{F}_c$ was made by restricting iterates in RP to the span of the first few singular functions of $\mathscr{F}_c$. When done for $c = 2$ the resulting iterates reduced $F(\hat{G}_n)$ at a slower rate, gave final iterates $\hat{G}^i$ for which $\|\hat{G}^i - G^i\|$ was approximately the same as in the unfiltered algorithm, and still failed to terminate due to the convergence of successive iterates but rather ended due to the satisfaction of one of the other two conditions. For this value of c there are approximately nineteen significantly nonzero singular values; the results indicate that even over this subspace the function $F(G)$ is widely varying due to the nonlinearity and existence of multiple solutions. It therefore seems likely that a severe restriction, e.g. to the span of the first five singular functions, is necessary to provide an easily solved problem.

The numerical results presented are some *typical* examples of the behavior of the algorithms on model problems 2 and 2a, with parameter pairs (c,N) of $(1.5,22)$, $(2,32)$ and $(2,40)$ and test function G, Eq. (6.1). For numerical

results concerning model problem 2b, see [15]. Tables 1 and 2 first give the average number of iterations $\tilde{n}$ and average final value $F(\hat{G}_{\tilde{n}})$ of RP when iterated to convergence on several different initial guesses, each perturbed by an amount ε_1 from the true solution. The remaining entries are the ratio $\breve{n}/\tilde{n}$, where $\breve{n}$ is the number of iterations required by the remaining algorithms to reduce $F(\hat{G}_{\breve{n}})$ to below $F(\hat{G}_n)$. Cases where the algorithm on trial failed to reduce $F(G_{\breve{n}})$ to within 100 $F(G_{\tilde{n}})$ are denoted by a *. Figures 4-7 show 4 typical final iterates reached by these algorithm for varying problems and values ε_1.

The filter parameter ε_3 of Eq. (5.14), (5.15), (5.22) and (5.23) at the n-th iteration was calculated from the quantity

$$\varepsilon = \max\{\min\{4\|\lambda_{n-1}H_{n-1}\| - .025, .2\}, .002\} \quad . \qquad (6.5)$$

For FN and FNP, ε_3 was taken to be ρ; for SQFN and SQFNP, $\varepsilon_3 = \rho/3$. This expression was calculated by trial and error and, although by no means the last word, at least has the property of giving search direction H_n such that the associated step length λ_n almost always lay in the interval [.1,1.5] and $\lambda_n \sim 1$ when close to the true solution.

7. CONCLUSION

The ill-posed nature of phase retrieval induced too much variation in numerical calculations to allow the drawing of quantitative judgements from these results, but some qualitative remarks can be made. For these one-dimensional problems RP was clearly the least expensive in terms of total computational cost to reach a desired objective, and it is difficult to see how any other algorithm can be improved through taking advantage of sparse Hessians, etc. to seriously challenge RP for this position. However SQFNP was the most robust in producing acceptable iterates from almost any starting point, and although it did not display to the same degree the apparent quadratic convergence of FN and FNP when close to the true solution, this could possibly be remedied by a better choice of filter.

Efforts to estimate the role of local ill-conditioning and filtering in phase retrieval problems were largely frustrated by the effects of global ill-conditioning due to the existence of multiple solutions. This makes one dimensional problems unsolvable for practical purposes, but it is expected that in higher dimensional problems of real interest, with fewer possible exact solutions, that local ill-conditioning will become dominant. Hopefully this may be removed by filtering and quadratically convergent algorithms will have a wider domain of convergence, thus becoming competitive with the present first order iterative schemes.

Table 1. Relative performance of algorithms on example 2 with test function G, Eq. (6.1).

ε_1	N			FN	FNP	SQFN	SQFNP
	22	1.9×10^{-5}	50	1.5	1.0	.40	.30
60	32	1.5×10^{-5}	20	*	.825	.875	.45
	40	3×10^{-5}	40	*	1.0	.35	.30
	22	5×10^{-5}	50	1.5	.30	.60	.30
100	32	4×10^{-4}	20	1.6	.45	.35	.40
	40	8×10^{-4}	40	.50	.50	.425	.30

Table 2. Relative performance of algorithms on example 2a with test function G, Eq. (6.1).

ε_1	N			FN	FNP	SQFN	SQFNP
20	22	9×10^{-6}	23	.20	.20	.50	.50
60	22	5×10^{-4}	50	.50	.30	.35	.45
	32	3×10^{-4}	20	2.0	.45	.50	.60
100	22	2×10^{-3}	50	3.0	.80	.325	.325
	32	6×10^{-2}	20	*	1.0	.50	.40

REFERENCES

1. R. Gerschberg and W. Saxton, "A practical algorithm for the determination of phase from image and diffraction plane pictures," Optik, 35, 237-246 (1972).

2. A. Papoulis, "A new algorithm in spectral analysis and bandlimited extrapolation," IEEE Trans. Circuits Syst., CAS22, 735-742 (1975).

3. D.C. Youla, "Generalized image restoration by the method of alternating orthogonal projections," IEEE Trans. Circuits Syst., CAS25, 695-702 (1978).

4. J.A. Cadzow, "An extrapolation procedure for bandlimited signals," IEEE Trans. Acoust. Speech Signal Process., ASSP-27, 4-12 (1979).

5. M.S. Sabri and W. Steenart, "An approach to bandlimited extrapolation: The extrapolation matrix," IEEE Trans. Circuits Syst., CAS25, 74-78 (1978).

6. R. Burge, M. Fiddy, A. Greenway, and G. Ross, "The phase problem," Proc. Roy. Soc., A350, 191-212 (1976).

7. J.M. Ortega and W.C. Rheinboldt, Iterative Solution of Nonlinear Equations in Several Variables (Academic Press, New York, 1970), pp. 494-500.

8. J.R. Fienup, "Space-object imaging through the atmosphere," Opt. Eng., 18, 529-534 (1979).

9. R. Barakat and G. Newsam, "A numerically stable iterative method for the inversion of wavefront aberrations from measured point spread function data," J. Opt. Soc. Am., 70, 1255-1263 (1980).

10. A.N. Tikhonov and V.Y. Arsenin, Solutions of Ill-Posed Probelsm (Halsted, New York, 1977).

11. L.S. Taylor, "The phase retrieval problem," IEEE Trans. Antennas Prop. AP39, 381-391 (1981).

12. D.L. Misell, "The phase problem in electron microscopy," in Advances in Optics and Electron Microscopy, Vol. 7, eds. V.E. Coslett and R. Barer (Academic Press, New York, 1978).

13. H.A. Ferwerda, "The phase reconstruction problem for wave amplitudes and coherence functions," in Inverse Source Problems in Optics, ed. H.P. Baltes (Springer-Verlag, New York, 1978).

14. R. Barakat and G. Newsam, "Algorithms for reconstruction of partially known, bandlimited Fourier transform pairs from noisy data: I the prototypical linear problem." Submitted for publication to J. Integral Eqs.

15. R. Barakat and G. Newsam, "Algorithms for reconstruction of partially known, bandlimited Fourier transform pairs from noisy data: II the nonlinear problem of phase retrieval." Submitted for publication to J. Integral Eqs.

16. G. Newsam, Numerical Reconstruction of Partially Known Transforms, Ph.D. Thesis (Harvard University, 1982).

17. C.T. Baker, The Numerical Treatment of Integral Equations (Oxford Univ. Press, Oxford, 1977).

18. G. Newsam and R. Barakat, "The essential dimension as a well defined number of degrees of freedom of finite convolution operators appearing in optics." Submitted for publication to Opt. Acta.

19. R.H. Hanson, "A numerical method for solving Fredholm integral equations of the first kind using singular values." SIAM J. Numer. Anal., 8, 616-622 (1971).

20. D. Slepian and H. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty: I." Bell Syst. Tech. J., 40, 43-64 (1961).

21. H.J. Landau and H. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty: II." Bell Syst. Tech. J., 40, 65-84 (1961).

22. H.J. Landau and H. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty: III." Bell Syst. Tech. J., 41, 1295-1336 (1962).

23. H.J. Landau, "The eigenvalue behavior of certain convolution equations." Trans. Amer. Math. Soc., 11S, 566-569 (1964).

24. H.J. Landau, "Necessary density conditions for sampling and interpolation of certain entire functions." Acta Math., 117, 37-52 (1967).

25. H.J. Landau and H. Widom, "Eigenvalue distribution of time and frequency limiting." J. Math. Anal. and Applics., 469-481 (1980).

26. H. Widom, "Asymptotic behavior of the eigenvalues of certain integral equations: II." Arch. Rat. Mech. Anal., 17, 215-229 (1964).

27. E.J. Akutowicz, "On the determination of the phase of a Fourier integral -I." Trans. Amer. Math. Soc., 83, 179-192 (1956).

28. E.J. Akutowicz, "On the determination of the phase of a Fourier integral -II." Proc. Amer. Math. Soc., 8, 234-238 (1957).

29. A. Walther, "The question of phase retrieval in optics." Opt. Acta, 10, 41-49 (1963).

30. E.M. Hofstetter, "Construction of time-limited functions with specified autocorrelation functions." IEEE Trans. Inf. Thy., IT-10, 119-126 (1964).

31. R. Barakat and G. Newsam, "On the existence of multiple solutions to the two-dimensional phase recovery problem." Submitted to Opt. Comm.

32. P.P. Boas, Entire Functions (Academic Press, New York, 1954).

33. P.J. Napier, "The brightness temperature distributions defined by a measured intensity interferogram." N.Z.J. Sci., 15, 342-355.

34. L. Ronkin, Introduction to the Theory of Entire Functions of Several Complex Variables (Amer. Math. Soc., Providence, 1974).

35. W. Lawton, "Uniqueness results for the phase retrieval problem for radial functions." J. Opt. Soc. Am., 71, 1519-1522 (1981).

36. J.R. Fienup, "Phase retrieval algorithms: a comparison." Appl. Opt., 21, 2758-2769 (1982).

37. L. Gubin, B. Polyak, and E. Raik, "The method of projections for finding the common point of convex sets," USSR Comp. Math. and Math. Phys., 7, 1-24 (1967).

38. Y. Censor and G.T. Herman, "Row generation methods for feasibility and optimization problems involving sparse matrices and their applications." In: Sparse Matrix Proceedings 1978, eds. I. Duff and G. Stewart (SIAM, Philadelphia, 1979).

39. G. Walsh, Methods of Optimization (Wiley, New York, 1975).

40. R.B. Holmes, "Smoothness of certain metric projections on Hilbert spaces." Trans. Amer. Math. Soc., 183, 87-100 (1973).

41. R. Barakat and L. Riseberg, "Diffraction theory of the aberrations of a slit aperture." J. Opt. Soc. Am., 55, 878-881 (1965).

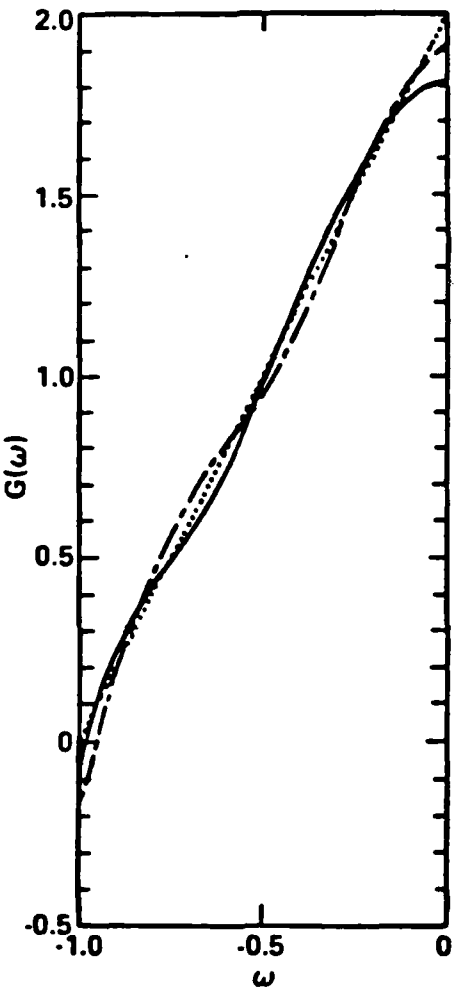

Fig. 1. Two sample realizations (— · — and —) of the reconstruction of $G(\omega)$, Eq. (2.16), in the presence of 3% noise in $\mathcal{F}_c G$, Eq. (2.14).

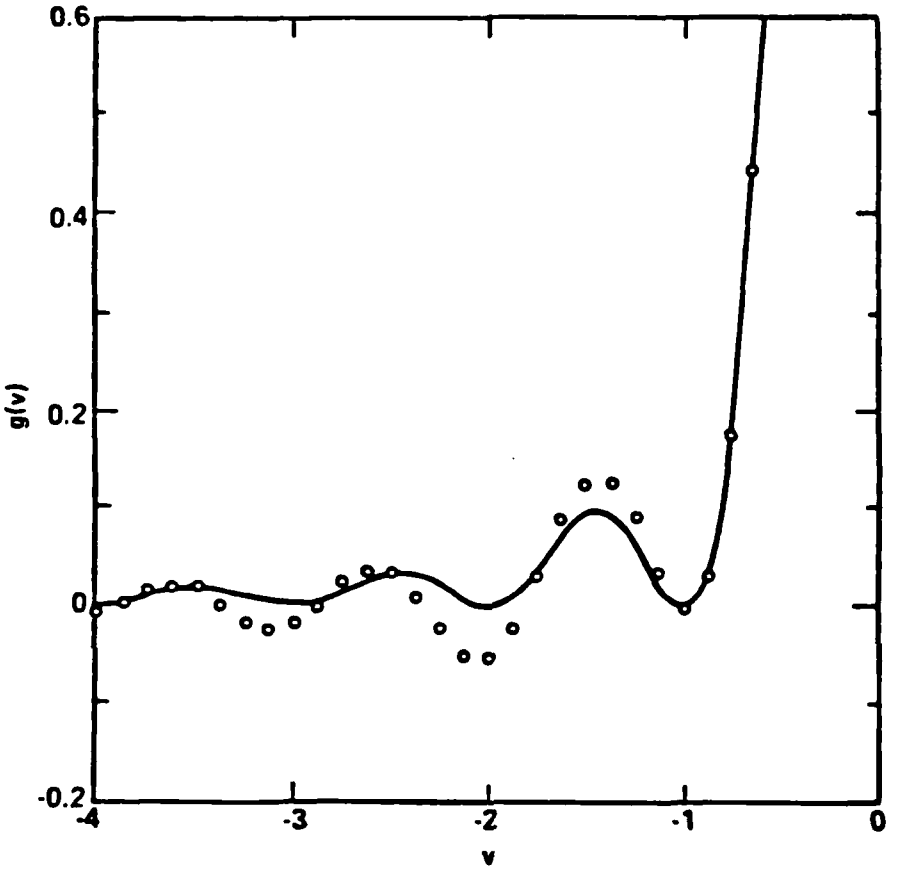

Fig. 2. Extrapolation of g(v), Eq. (2.15), see open circles, corresponding to — · — in Fig. 1.

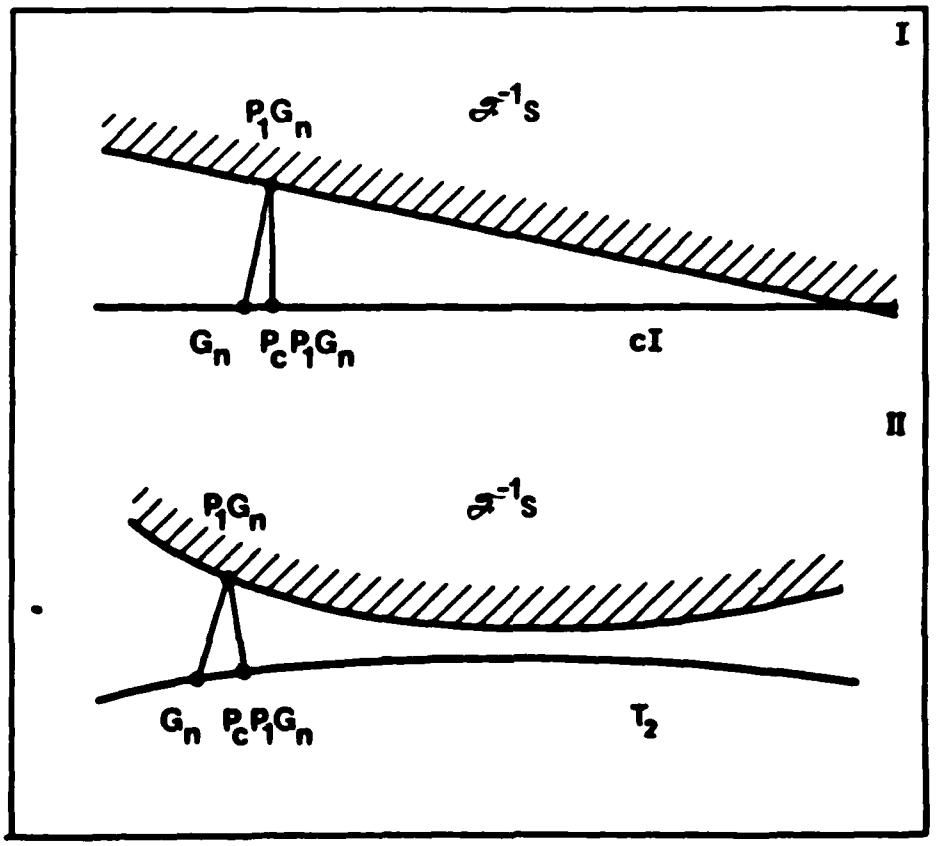

Fig. 3. Two examples of poor convergence of the alternating projection algorithm induced by ill-conditioning.

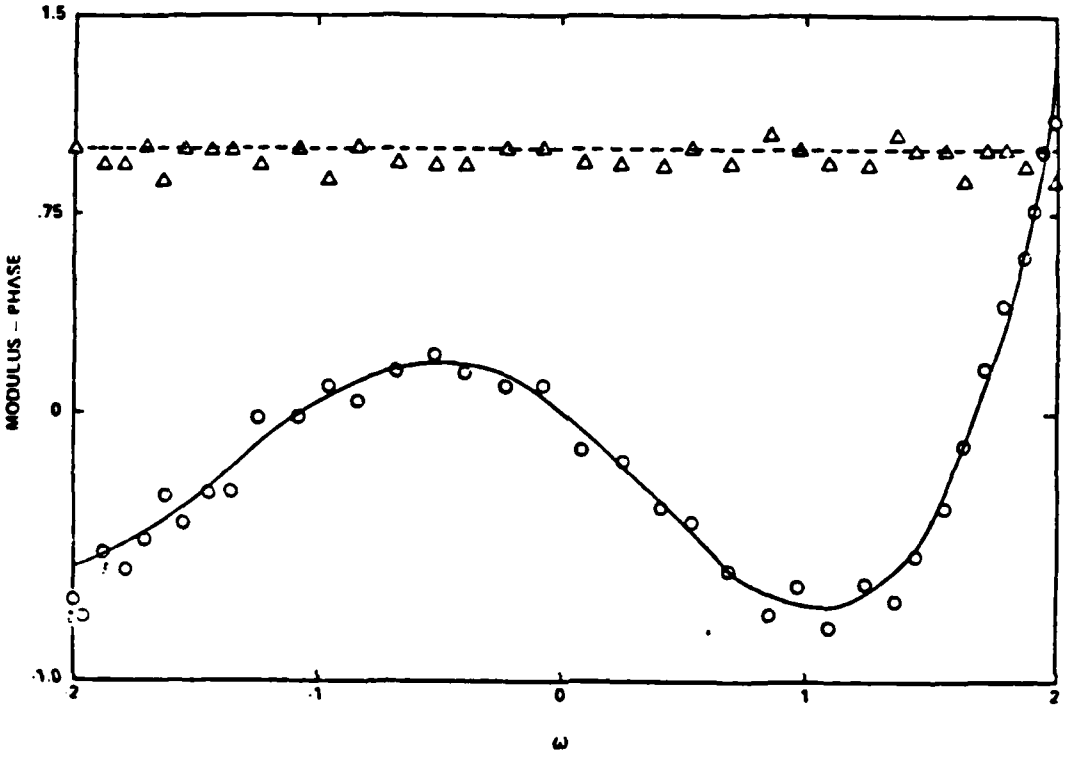

Fig. 4. Solution $\hat{G}$ to example 2 reached from an initial guess $\hat{G}_0$ with $\varepsilon_1 = 20\%$: —— exact phase from Eq. (6.1); o reconstructed phase; ---- exact modulus from Eq. (6.1); Δ reconstructed modulus.

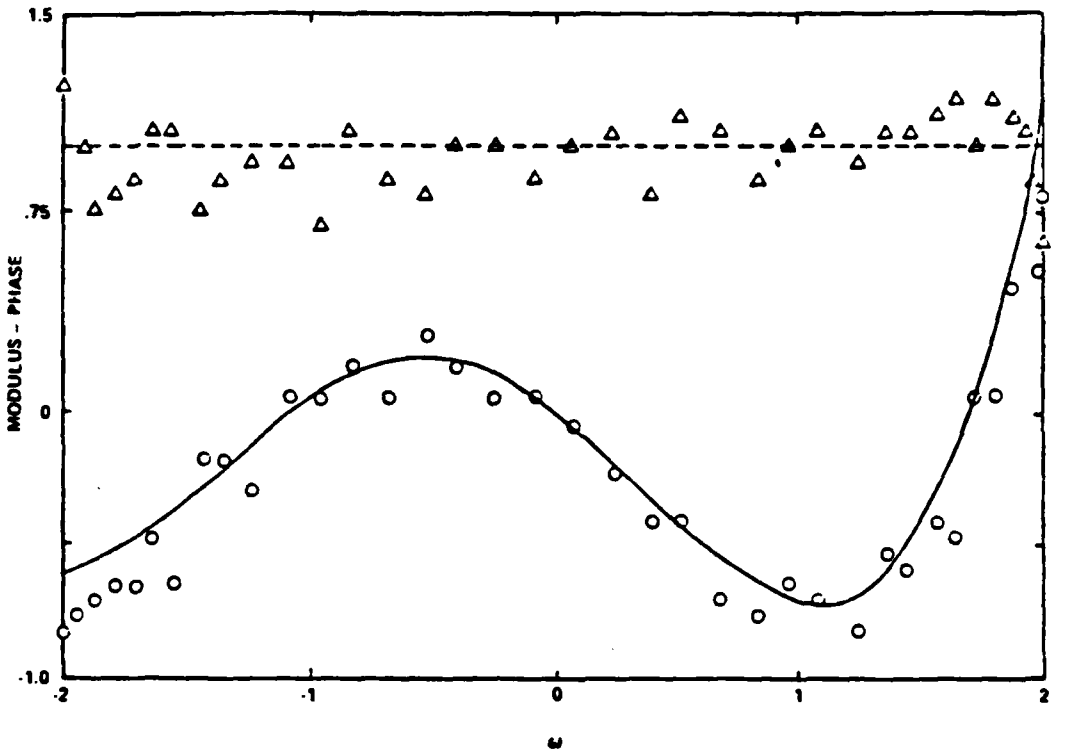

Fig. 5. Solution $\hat{G}$ to example 2 reached from an initial guess $\hat{G}_0$ with $\varepsilon_1 = 60\%$: —— exact phase from Eq. (6.1); o reconstructed phase; ---- exact modulus from Eq. (6.1); Δ reconstructed modulus.

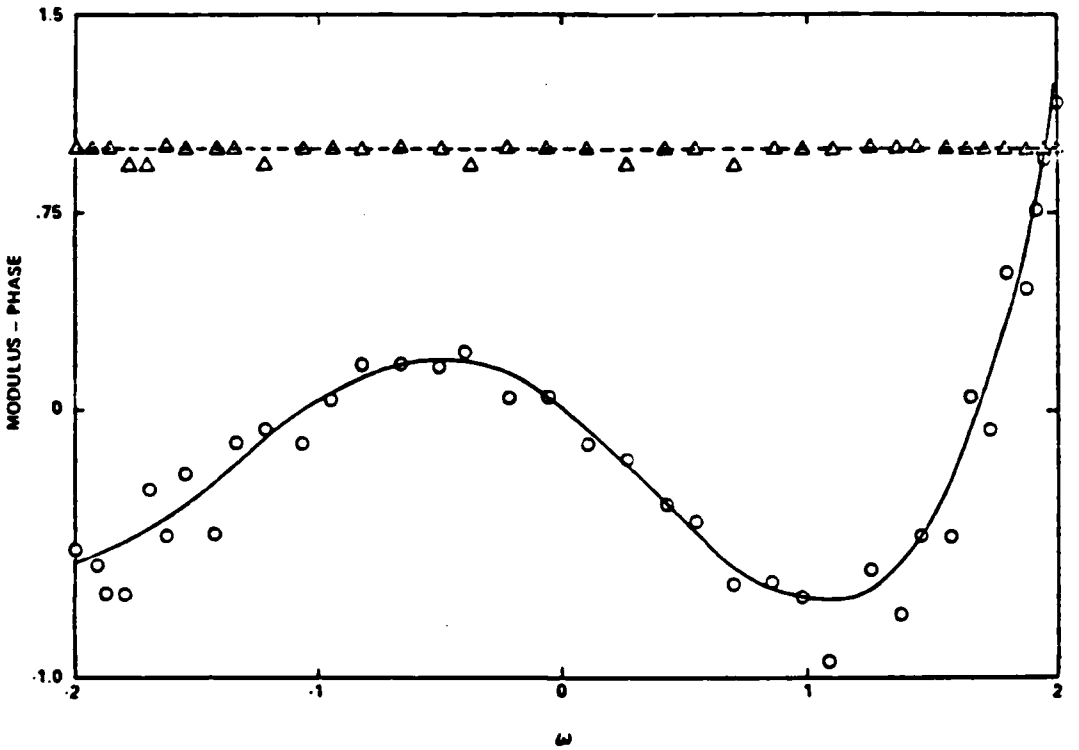

Fig. 6. Solution $\hat{G}$ to example 2a reached from an initial guess $\hat{G}_0$ with $\varepsilon_1 = 60\%$: —— exact phase from Eq. (6.1); o reconstructed phase; ---- exact modulus from Eq. (6.1); Δ reconstructed modulus.

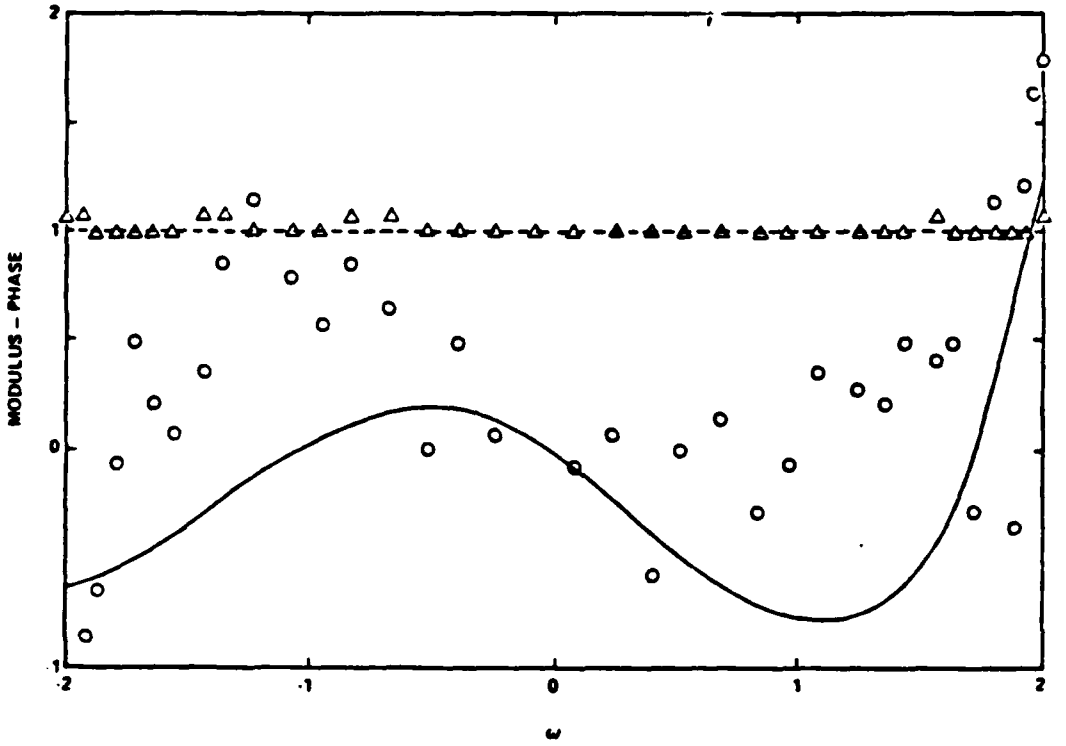

Fig. 7. Solution $\hat{G}$ to example 2a reached from an initial guess $\hat{G}_0$ with $\varepsilon_1 = 100\%$: —— exact phase from Eq. (6.1); o reconstructed phase; ---- exact modulus from Eq. (6.1); Δ reconstructed modulus.

ALGORITHMS FOR RECONSTRUCTION OF PARTIALLY KNOWN, BANDLIMITED FOURIER TRANSFORM PAIRS FROM NOISY DATA: I THE PROTOTYPICAL LINEAR PROBLEM

ABSTRACT

Problems involving reconstruction of partially known, bandlimited Fourier transform pairs from noisy data are now regularly encountered in a wide variety of scientific and technical areas. This paper is the first in a series of studies of algorithms for their solution. These studies focus on the algorithmic structure with respect to the dominant feature of such problems, that they are ill-posed. The algorithms developed here are for the linear prototype problem; namely extrapolation of a bandlimited function known over a finite interval. The problem is cast as the inversion of a linear operator, the finite Fourier transform. Criteria can be deduced from a knowledge of the spectrum of this operator for the suitability of extrapolation algorithms. These criteria are used to evaluate existing algorithms (such as the Gerschberg-Saxton) as well as our new algorithm based upon inversion of a Galerkin approximation to the operator using singular value decomposition. Numerical results on the relative merits of different discretizations in our algorithm, and on its success in extrapolation in two examples (with optical diffraction interpretations) with noisy data are presented.

1. INTRODUCTION

A recurring problem in many fields is the reconstruction of a Fourier transform pair g, G from partial data on either or both functions. Considerable effort has been expended in the development of algorithms for its solution; although there have been some successes, the problem has generally proved to be difficult. In a series of papers, of which this is the first, we aim to detail the features that cause difficulty for numerical calculations and to construct algorithms that take explicit account of such features. It is not possible to remove such difficulties, but if they are ignored in the construction of algorithms they invariably surface later in the solution in a more inconvenient form.

The canonical examples for such reconstructions are:

Example 1, Extrapolation of Band-Limited Signals. Given a noisy measurement of g on an interval $[a_1,a_2]$ and the knowledge that G vanishes outside the bounded interval $[b_1,b_2]$, reconstruct g and G on the entire real line.

Example 1 is the archetypical linear problem in transform reconstruction. A number of algorithms have been proposed for its solution, either by iterative means: Gerschberg and Saxton [1], Papoulis [2], Youla [3] or by direct means: Cadzow [4], Sabri and Steenaart [5].

Example 2, The Phase Problem. Given a noisy measurement of $|g|$ on an interval $[a_1,a_2]$ and the knowledge that G vanishes outside the bounded interval $[b_1,b_2]$, reconstruct g and G on the entire real line.

Example 2 has been of theoretical interest for some time; see for example Burge, Fiddy, Greenway, and Ross [6]; however in this most general form it has proved intractable. Numerical solutions obtained in particular cases have done so by (sensibly) incorporating further knowledge of g and G. Two such examples are:

Example 2a, The Two Moduli Problem. Given noisy measurements of $|g|$ on $[a_1,a_2]$, $|G|$ on $[b_1,b_2]$, and the knowledge that G vanishes identically outside of $[b_1,b_2]$, reconstruct g, G over the entire real line.

Example 2b, The Phase Problem with Nonnegativity Constraints. Given noisy measurements of $|g|$ on $[a_1,a_2]$ and the knowledge that G is nonnegative over $[b_1,b_2]$ and vanishes identically outside $[b_1,b_2]$, reconstruct g, G over the real line.

Gerschberg and Saxton [1] developed a widely used algorithm for Example 2a; almost all the iterative algorithms for the solution of the general reconstruction problem are suitably modified versions of this particular case. The Gerschberg-Saxton algorithm (hereafter denoted as the GS algorithm) in its general form is essentially the steepest descent algorithm with unit step length and so is first order [7]. An alternative second order algorithm, based on Newton's method, has been proposed by Barakat and Newsam [8]. In [9], the GS algorithm is successfully applied to a problem of the type in Example 2b.

The canonical examples presented are simple in form, nevertheless they contain the salient features that make other, more complicated, problems intractable. The font of all difficulties is that the reconstruction problem is ill posed, and badly ill posed at that. The original definition of a well posed problem is due to Hadamard [10].

Definition. A problem is well posed if the solution

1. exists
2. is unique
3. depends continuously on the data.

If a problem violates any of these conditions, it is ill posed.

Example 1 satisfies only condition 2; therefore it is ill posed. To see that it fails condition 1, it suffices to note that g is the transform of a function with bounded support so that it is analytic by the Paley-Wiener theorem [11]. Any perturbation of the true g by noise that is not analytic produces a perturbed problem with no solution. Analyticity does imply condition 2, but also that condition 3 will fail very badly. This is due to the natural error metric for noise being the standard energy norm, which is of no use as a measure of analyticity. Therefore noise that is arbitrarily small in the error metric will produce arbitrarily large changes in the solution, or even cause it not to exist. This aspect of the problem is discussed in section 2.

Example 2 fails all three conditions; condition 1 for the same reason as Example 2. Akutowicz [12,13] shows the existence of multiple solutions for the phase problem both with and without nonnegativity constraints. It is not known whether the two moduli problem has multiple solutions, except for some trivial multiplicities due to symmetry and constant phase factors. As for condition 3, Example 2 can be locally linearized into Example 1, so it also fails this condition.

Many researchers do not seem to be aware of the full implications of these failures for numerical algorithms for approximate solutions. Gerschberg and Saxton have been careful to show that their algorithm is always error decreasing and to give examples of behavior using noisy data. Chapman [14] modified their algorithm to make even greater use of known noise levels. However the overall approach is still on an *ad hoc* basis, noise cannot be explicitly handled in a satisfactory fashion in an iterative algorithm. Papoulis [2] gives error bounds for his iterative method, but not a way of incorporating them efficiently into calculations. It will be shown in this paper that direct methods do allow development of robust algorithms. The two direct methods proposed in Cadzow [3], Sabri and Steenaart [4] call for inversion of ill conditioned matrices formed with no concern for optimizing the condition number. The optics literature contains many references to "the number of degrees of freedom" [15,16].

However, the natural corollary has not been noted: numerical solutions to this problem should contain discretizations of approximately this size and no significant improvement in results can be obtained by using greater numbers of points, approximating functions, etc. Ignorance of this point has lead to unnecessarily long calculations in the belief that more points give greater accuracy.

This paper aims to provide a detailed study of Example 1 in the language of numerical analysis, and to use this knowledge to construct numerical algorithms for its solution in a rational manner. Section 2 contains an exposition of the structure of the finite Fourier transform and the linear operator associated with Example 1. Using this structure in conjunction with the theory of singular value decomposition, the solution is decomposed into

two parts; the first of which is finite dimensional and contains significant information in the presence of noise,while the second contains no trustworthy information in the presence of noise. In section 3, numerical algorithms are proposed that efficiently identify these two components, in addition a list of desirable features of possible discretizations is given. Section 4 contains an exposition of iterative methods, and a comparison of their merits with those of the direct methods of section 3. Finally section 5 contains a discussion of some particular discretizations with numerical calculations showing their robustness in the presence of noise.

A clear exposition of the linear problem is important not only in its own right, because as mentioned above the local structure of nonlinear reconstructions can be approximated by linearization in the neighborhood. The results of section 2 on this linearization show that the nonlinear reconstructions are essentially problems over finite dimensional manifolds. If Newton's method is used to carry out these reconstructions, an accurate efficient way of identifying the tangent plane to this manifold is needed. Reference [8] shows that singular value decomposition can accomplish this task. We believe that the results of this paper show that Newton's method coupled with careful discretizations for nonlinear reconstructions can be competitive with the widely used iterative methods.

2. THEORY OF THE MODEL LINEAR PROBLEM

This section contains an analysis of the theory of the linear reconstruction problem discussed in the previous section. A formal description is:

Given a function $\hat{g}(v)$ measured on $v \in A \equiv [a_1, a_2]$ that is known to be in the Fourier transform of a function $G(\omega)$ which has support contained in $B \equiv [b_1, b_2]$, extend $\hat{g}(v)$ to a function $g(v)$ defined on the entire real axis.

For convenience we introduce the following notation.

Definition 1. Let $G(\omega) \in L^2$, the Fourier transform $g(v)$ of $G(\omega)$ is defined as

$$g(v) \equiv \int_{-\infty}^{\infty} e^{2\pi i v\omega} G(\omega) d\omega \tag{2.1}$$

and denoted by

$$g = \mathcal{F}G \tag{2.2}$$

The inverse transform is denoted by

$$G = \mathcal{F}^{-1} g \tag{2.3}$$

Definition 2. Let S be a subset of the real line. The projection operator on L^2 associated with S is denoted by P_s and defined by

$$\begin{aligned} (P_s f)(v) &= f(v) \qquad v \in S \\ &= 0 \qquad\quad v \notin S \end{aligned} \tag{2.4}$$

In this notation, the model problem is: Given a measurement $\hat{g}$, find g and G such that

$$\hat{g} = P_B \mathcal{F} P_A G \quad , \quad g = \mathcal{F} P_A G \tag{2.5}$$

The problem can be symmetrized using simple translations and scaling to read: Given a measurement $\hat{g}$, find g, G such that

$$\hat{g} = P_c \mathcal{F} P_c G \quad , \quad g = \mathcal{F} P_c G \tag{2.6}$$

where

$$C = [-c,c] \quad , \quad c \equiv \frac{1}{2}[(a_2 - a_1)(b_2 - b_1)]^{1/2} \tag{2.7}$$

To solve Equation 2.6, we need to convert the linear operator $P_c \mathcal{F} P_c$. As noted in the previous section this is an ill posed problem. By the Paley-Wiener theorem [11]; $\hat{g}(v)$, the Fourier transform of a function with bounded support is analytic; consequently it has a unique extension $g(v)$. Therefore the model problem is one of analytic continuation which is known to be highly sensitive to measurement error in $\hat{g}(v)$, see [1,8]. In order to make quantitative judgements on the effects of noise in Equation 2.6, we shall use the concept of singular value decomposition. We assume the reader to be familiar with this technique applied to either infinite dimensional compact linear operators or to finite dimensional matrices; a discussion of the first case can be found in Baker [17], of the second in Stewart [18].

In a series of papers, Slepian et al. [19-21], the structure of $P_c \mathcal{F} P_c$ was fully explored. The key results are repeated here:

1) $P_c \mathcal{F} P_c$ is a normal compact operator, and

$$\| P_c \mathcal{F} P_c \|_2 \leq 1 \tag{2.8}$$

2) The eigenfunctions of $P_c \mathcal{F} P_c$ can be taken as real. They are then the prolate spheriodal wavefunctions here denoted by $\phi_n(v,c)$, making explicit the dependence on c.

3) The eigenvalues $\lambda_n(c)$ satisfy

$$\lambda_n(c) = e^{i\pi n/2} |\lambda_n(c)| \tag{2.9}$$

$$|\lambda_n| > |\lambda_{n+1}| > 0$$

4) The singular values are denoted by $\sigma_n(c)$ with

$$\sigma_n(c) \equiv |\lambda_n(c)| \tag{2.10}$$

They behave as follows

a. For a fixed c and increasing n

$$\sigma_n(c) \sim (c^2/n)^{2n} \tag{2.11}$$

b. For a fixed n, there exist constants α_n, β_n such that for increasing c

$$\sigma_n(c) \sim 1 - \alpha_n e^{-\beta_n c^2} \tag{2.12}$$

Since $P_c \mathcal{F} P_c$ is compact, it has a singular value decomposition U, Σ, V where U and V are complete sets of orthonormal functions, and Σ is the decreasing sequence of nonnegative numbers defined in Equation 2.10. Since $P_c \mathcal{F} P_c$ is normal

$$U = \{u_n \phi_n(\omega,c)\}_{n=0}^{\infty}$$

$$V = \{v_n \phi_n(\omega,c)\}_{n=0}^{\infty} \tag{2.13}$$

where u_n and v_n are constants such that $|u_n| = |v_n| = 1$. With this machinery, the solution to the problem can theoretically be determined by expanding $\hat{g}$ and G as

$$\hat{g}(v) = \sum_{n=0}^{\infty} a_n v_n \phi_n(v,c) \tag{2.14}$$

$$G(\omega) = \sum_{n=0}^{\infty} b_n u_n \phi_n(\omega,c) \tag{2.15}$$

The coefficients a_n are determined directly from the known $\hat{g}$ and the b_n from the relation

$$\hat{g} = P_c \mathcal{F} P_c G \tag{2.16}$$

the final result is

$$b_n = \frac{a_n}{\sigma_n} \tag{2.17}$$

In this representation, the degree of ill conditioning of Equation 2.6 (i.e. the measure of how ill posed Equation 2.6 is) as well as the effect of noise can be determined from Equation 2.17 and the behavior of σ_n. Intuitively since $\sigma_n \to 0$, by Equation 2.12 a small error in the coefficients a_n of a high frequency (large n) component $a_n v_n \phi_n(v,c)$ of $g(v)$ will induce a large error in the coefficient b_n, given by Equation 2.17, of the corresponding component of $G(\omega)$. A more precise statement is

Lemma 1 Suppose the relative error (in the L^2 norm) on the measured g is ε, and the maximum absolute error that will be tolerated in G is δ. Then there is a decomposition of G into two components G_1 and G_2, in which G_1 is always accurate within an error δ and G_2 cannot be guaranteed to be this accurate.

Proof: Let G be decomposed as the sum in Equation 2.15. For a fixed N, the worst possible absolute error in the first N components of G is induced by assuming that $a_1,\ldots,a_{N-1}$ are precisely known and that $a_1 = 1$, $a_2 = \cdots = a_{N-1} = 0$, a_N is in error by ε and $a_i = 0$ for $i > N$. Then

$$b_1 = 1/\sigma_1 \quad , \quad b_2 = \cdots = b_{N-1} = 0 \quad , \quad b_i = 0 \quad \text{for} \quad i > N$$

$$b_N = \varepsilon/\sigma_N = b_1(\varepsilon\sigma_1/\sigma_N) \tag{2.18}$$

thus the absolute error induced is $(\varepsilon\sigma_1/\sigma_N)$. If N is chosen so that

$$(\varepsilon\sigma_1/\sigma_N) < \delta \leq (\varepsilon\sigma_1/\sigma_{N+1}) \tag{2.19}$$

then the first N componenets of G are always known to within an absolute error δ, but the remaining components are not that trustworthy.

But as the decomposition allows a direct observation of the effects of noise, so also does it allow direct action to reduce these effects. A filter $f(\sigma,\varepsilon)$ dependent on the noise levels ε, δ is introduced so that G is estimated by G'

$$G'(\omega) = \sum_{n=0}^{\infty} a_n f(\sigma_n)\phi_n(\omega,c) \tag{2.20}$$

where

$$f(\sigma_n) \to (\sigma_n)^{-1} \quad \text{as} \quad \sigma_n \to 1$$

$$\to 0 \qquad \text{as} \quad \sigma_n \to 0 \tag{2.21}$$

In order to construct a sensible filter for the general problem, we first note the following conclusions from Equations 2.8 - 2.12.

a. $P_c \mathcal{F} P_c$ is badly ill conditioned because $\sigma_n(c)$ exhibits exponential decay in n

b. For a given c, the singular values are approximately distributed as a step function in that there exists an $N(c)$ such that

$$n < N(c) \Rightarrow \sigma_n(c) \sim 1$$

$$n > N(c) \Rightarrow \sigma_n(c) \sim 0 \tag{2.22}$$

c. For a given c and noise levels ε, δ let $N(\varepsilon,\delta,c)$ be the N of Equation 2.19. Then $N(\varepsilon,\delta,c)$ is bounded by Equation 2.12. Furthermore from conclusions a and b, $N(\varepsilon,\delta,c)$ is relatively insensitive to ε and δ (i.e. for almost all ε and δ there exists a $p << 1$ such that

$$N(\varepsilon,\delta,c)-1 \leq N(\varepsilon,p\delta,c) \leq N(\varepsilon,\delta,c)$$

$$\leq N(p\varepsilon,\delta,c) \leq N(\varepsilon,\delta,c) + 1 \tag{2.23}$$

The general choice of filters is an art, however here we make the particular choice

$$f(\sigma,\varepsilon) = 1/\sigma \quad , \quad \sigma > \varepsilon/\delta$$
$$= 0 \quad , \quad \sigma < \varepsilon/\delta \tag{2.24}$$

relying on conclusions b and c above. This filter effectively truncates the series in Equation 2.20 leaving only the trustworthy component G, of Lemma 1. Since the filter is insensitive to ε and δ, the number of elements in the truncated sum of Equation 2.20 is insensitive to noise levels It corresponds to the "essential number of degrees of freedom" of the model problem mentioned in section 1. We stress that this number is relatively noise independent (i.e. a function of c with weak dependence on ε and δ) and that it represents the number of components of the solution that are of guaranteed accuracy.

Based on these results, the following algorithm is proposed. For a given c, ε, and δ calculate $N(c,\varepsilon,\delta)$. If it is sufficiently small, accurate numerical approximations to σ_n, $\phi_n(v,c)$ are calculated for $n \leq N(c,\varepsilon,\delta)$ and G is estimated using Equation 2.20. This procedure is optimal in that it focuses on accurately calculating components of $\hat{g}$ that are significant in estimating G while ignoring components that are useless for estimation.

3. DISCRETIZATION AND ASSOCIATED THEORY

In this section we consider a general class of finite dimensional approximations to g, G and weigh the merits of different approximations. The discretizations considered are derived using Galerkin's method, they are based upon the natural error metric $\|\cdot\|_2$ for the model problem. We choose two sets of linearly independent functions on $L^2[-c,c]$

$$\{\psi_k\}_{k=1}^K \quad , \quad \{\theta_\ell\}_{\ell=1}^L \quad ; \quad \|\psi_k\|_2 = \|\theta_\ell\|_2 = 1 \tag{3.1}$$

and introduce

Definition 3. Let S_K denote the subspace spanned by $\{\psi_k\}_{k=1}^K$, S_L the subspace spanned by $\{\theta_\ell\}_{\ell=1}^L$; let P_K denote the projection operator from L^2 onto S_K, Q_L the projection operator from L^2 on S_L.

We generate approximations

$$\hat{g}_K \in S_K \to \hat{g} \quad , \quad G_L \in S_L \to G \tag{3.2}$$

by requiring that $\hat{g}_K$, G_L minimize

$$\|\hat{g} - h\|_2 \quad , \quad h \in S_K \tag{3.3}$$

$$\|\hat{g} - P_K P_C \mathcal{F} P_C H\|_2 \quad , \quad H \in S_L \tag{3.4}$$

Since S_K, S_L are finite dimensional, such approximations always exist. By expanding $\hat{g}_K$, G_L in terms of the basis functions

$$\hat{g}_K = \sum_{k=1}^K b_k \psi_k \quad , \quad G_L = \sum_{\ell=1}^L a_\ell \theta_\ell \tag{3.5}$$

and defining the quantities

$$c_k = (\psi_k , \hat{g}) \qquad , \qquad B_{ij} = (\psi_i , \psi_j) \tag{3.6}$$

$$A_{ij} = (\theta_i , \theta_j) \qquad , \qquad D_{k\ell} = (\psi_k, P_c \mathcal{F} P_c \theta_\ell) \tag{3.7}$$

then Equations 3.3 and 3.4 can be rewritten as discrete equations for the vectors $\tilde{a} \equiv (a_1,\ldots,a_L)^+$, $\tilde{b} = (b_1,\ldots,b_K)^+$.

We note that

$$\hat{g}_K \text{ minimizes Equation 3.3} \Leftrightarrow \tilde{B}\tilde{b} = \tilde{c} \tag{3.7}$$

$$G_L \text{ minimizes Equation 3.4} \Leftrightarrow \tilde{a} \text{ minimizes}$$

$$\| \tilde{B}^{-\frac{1}{2}} (\tilde{c} - \tilde{D}\,\tilde{d}) \| \qquad \tilde{d} \in R^L \tag{3.8}$$

where the matrices $\tilde{B}$, $\tilde{D}$ are defined in Equations 3.6 and 3.7. For typographic convenience, the subscript 2 will be omitted from $\|\cdot\|$.

In this construction, there is considerable freedom of choice for the subspaces S_K , S_L and their bases; it is therefore appropriate to adopt criteria by which a particular choice can be judged. We list six criteria for evaluation of approximations and for each criterion list features or bases that produce an approximation giving good results. We can then seek a "best fit" approximation to these features.

1. Uniqueness of approxmation: Because $\|\cdot\|$ is a strictly convex norm $\hat{g}_K$ is unique. If $K < L$ then G_L is not unique; if $K \geq L$ then G_L is unique unless $P_c \mathcal{F} P_c S_L$ contains components orthogonal to S_K.

Alignments of this form are pathological; the practical problem is the occurrence of *nearly* orthogonal components. These components will contribute to ill conditioning in $\tilde{D}$, however both the pathology and the ill conditioning can be avoided by choosing ψ_k so that

$$P_c \mathcal{F} P_c S_L \subseteq S_K \tag{3.9}$$

Because the true solution is unique there appears to be no gain in constructing nonunique approximations. Henceforth, we assume that $K \geq L$ and that G_L is unique.

2. *Accuracy of approximations*: An obvious requirement for any sequence of approximations $\hat{g}_K$, G_L is that

$$\lim_{K\to\infty} \|\hat{g} - \hat{g}_K\| = 0 \tag{3.10}$$

and if g is noiseless that

$$\lim_{K,L\to\infty} \|G - G_L\| = 0 \tag{3.11}$$

Necessary conditions for Equations 3.10 and 3.11 to hold are that the sequences $\{\psi_k\}_{k=1}^{\infty}$, $\{\theta_\ell\}_{=1}^{\infty}$ be complete sets in $L^2[-c,c]$. These conditions are also sufficient for Equation 3.10, but not for Equation 3.11 because the inverse of $P_c \mathcal{F} P_c$ is unbounded

The important issue for numerical calculations is the rate of convergence of such approximations. Rather than attempt direct estimates of $\|G - G_L\|$ we consider a more easily computed quantity. The approximating subspaces chosen implicitly define an associated finite dimensional operator

$$R_{KL} = P_K P_c \mathscr{F} P_c Q_L \tag{3.12}$$

The quantity $\|R_{KL} - P_c \mathscr{F} P_c\|$ can then serve as a measure of the accuracy of the approximation.

We therefore seek functions ψ_k , θ_ℓ such that

$$\hat{g}_K \ , \ G_L \to \hat{g} \ , \ G \tag{3.13}$$

and that minimize this error metric. The following proposition provides a partial answer

Proposition 1: If $K = L = N$ and ψ_k , θ_ℓ are chosen so that $\psi_k(v) = \theta_k(v) = \phi_k(v)$ (the k-th eigenfunction of $P_c \mathscr{F} P_c$), then the Galerkin approximation generated by Equations 3.3, 3.4 satisfy Equations 3.10, 3.11 as $N \to \infty$. Furthermore for any other subspces $S_{K'}$, $S_{L'}$, with K' , $L' \leq N$,

$$\|R_{NN} - P_c \mathscr{F} P_c\| \leq \|R_{K'L'} - P_c \mathscr{F} P_c\| \tag{3.14}$$

The proof of this and the succeeding two propositions can also be found in [ ]. Thus the eigenfunctions are an "optimal" basis for construction of approximations under this criterion

3. *Conditioning of* $\tilde{A}$, $\tilde{B}$, $\tilde{D}$: The previous criterion aids in distinguishing suitable approximating subspaces S_K , S_L . One must now make a choice of bases within these subspaces. It is possible to choose ψ_k , θ_ℓ so that $\tilde{D}$ is well conditioned, but this is an illusory gain because the matrices $\tilde{B}$ and/or $\tilde{A}$ will then be ill conditioned. Therefore calculations of $\hat{g}_K(v)$, $G_L(\omega)$ made by using the representations in Equation 3.5 will be

error prone. Since the most informative measure of ill conditioning in a metric, the condition number, is not easily calculated; we use a less informetive, but more easily manipulated, measure for comparison of different bases, the determinant.

Proposition 2: Let S_K, S_L be fixed subspaces with orthonormal bases $\{\psi_k\}_{k=1}^K$, $\{\theta_\ell\}_{\ell=1}^L$. If $\{\psi_k'\}_{k=1}^K$, $\{\theta_\ell'\}_{\ell=1}^L$ are any other bases for S_K, S_L and the corresponding matrices $\tilde{A}$, $\tilde{A}'$, $\tilde{B}$, $\tilde{B}'$, $\tilde{D}$, $\tilde{D}'$ are defined by Equations 3.5, then

$$|\det(\tilde{B}'\tilde{D}'\tilde{A}')| \leq |\det(\tilde{B}\tilde{D}\tilde{A})| = |\det(\tilde{D})| \quad . \tag{3.15}$$

Proposition 3: Let $K = L = N$, $\psi_k = \theta_k = \phi_k$. Then for any other subspaces $S_{K'}$, $S_{L'}$ with $K', L' \geq N$

$$|\det \tilde{D}'| \leq |\det \tilde{D}| \tag{3.16}$$

These two propositions demonstrate that under this measure of ill conditioning, orthonormal bases should be chosen and that the best choice of bases is the eigenfunctions.

4. *Desirable special results:* Particular choices of bases yield interesting additional results. If both bases are orthonormal, then the spectral properties of $\tilde{D}$ are identical with those of R_{KL}, which in turn approximate those of $P_c \mathcal{F} P_c$. If $K = L$ and ψ_k, θ_k satisfy

$$\psi_k = P_c \mathcal{F} P_c \, \theta_k \tag{3.17}$$

then the Galerkin approximation is also a least squares solution, i.e. G_L minimizes

$$\|\hat{g} - P_c \mathcal{F} P_c H\| \qquad H \in S_L \tag{3.18}$$

Both these results are obtained if $K = L = N$, $\psi_k = \theta_k = \phi_k$.

5. *Presence of noise:* The above criterion are concerned only with properties of the approximations. The important property of the data, that it contains noise, must be considered. In section 2 we showed that it forced a division of the solution into trustworthy and untrustworthy components. The approximations therefore must echo this division as accurately as possible.

6. *Ease of computation:* In light of the above discussion the obvious choice for a basis are the eigenfunctions, but we do not consider their use as there is no known simple numerical algorithm for their calculation. To keep computing costs down ψ_k, θ_ℓ must have a simple closed form representation. Furthermore the functions $\mathcal{F} P_c \theta_\ell$ should also be available in a closed form so that calculation of an approximate interpolation g_L to g

$$g_L = \mathcal{F} P_c G_L \tag{3.19}$$

is cheap compared to numerical evaluation. If possible the inner products $(\psi_k, P_c \mathcal{F} P_c \theta_\ell)$ should also be found explicitly.

Taking all these issues into consideration we seek basis of simply calculated functions $\{\rho_k\}_{k=1}^{\infty}$ which are approximations to the eigenfunctions ϕ_k ; or at least possess, to some degree, the properties discussed in the criteria above. In particular the basis should be complete, orthonormal, and have the property that each eigenfunction can be expressed as a rapidly convergent series in ρ_k.

Given such a basis, we now describe an algorithm that for a given observation $\hat{g}$ and noise levels ε, δ produces a Galerkin approximation G_N to a component of G that is accurate up to an error δ, as in Lemma 1.

1. N, $K(N)$ and the subspace $S_{K(N)}$ are defined by

   a. N is the $N(\varepsilon,\delta,c)$ of Equation 2.19

   b. $S_{K(N)} = \text{span } \{\rho_k\}_{k=1}^{K(N)}$

   c. $K(N)$ is the least integer such that the finite dimensional operator R_{KL} associated with $K = L = K(N)$, $S_K = S_L = S_{K(N)}$ satisfies

$$\| \Gamma_{R_{K(N)K(N)}} - P_C \mathcal{F} P_C \| < \min(\sigma_N - \frac{\varepsilon}{\delta}, \varepsilon) \tag{3.20}$$

2. The $\tilde{c}$, $\tilde{D}$ of Equation 3.5 are formed for this choice of basis and the singular value decomposition [18]

$$\tilde{D} = \tilde{U} \tilde{\Sigma} \tilde{V}^{+} \tag{3.21}$$

calculated.

3. The subspaces $S_K = S_L = S_N$ from which the Galerkin approximation G_N will be formed by Equation 3.4 are the subspaces of $S_{K(N)}$, being the span of the first N eigenfunctions of $R_{K(N)K(N)}$, i.e.

$$\psi_k(v) = \theta_k(v) = \sum_{i=1}^{K(N)} V_{ik} \rho_i(v) \quad , \quad k = 1,\ldots,N \tag{3.22}$$

4. The approximant $\hat{g}_N$ to $\hat{g}$ from S_N can be formed from the approxmation $\hat{g}_{K(N)}$ to $\hat{g}$ in $S_{K(N)}$ (i.e. the vector $\tilde{c}$). Moreover G_N can be formed directly from $\tilde{c}$ as

$$G_N(\omega) = \sum_{k=1}^{K(N)} a_k \rho_k(\omega) \tag{3.23}$$

with

$$\tilde{a} = \tilde{V} \tilde{\Sigma}^+ \tilde{U}^+ \tilde{c}$$

where

$$\Sigma^+_{k\ell} = \Sigma^{-1}_{k\ell} \quad , \quad \text{if} \quad \Sigma_{k\ell} > \frac{\varepsilon}{\delta} \tag{3.24}$$

$$= 0 \quad , \quad \text{otherwise}$$

5. The approximate extrapolation g_N to g is now defined as

$$g_N(v) = \sum_{k=1}^{K(N)} a_k \mathcal{F} P_c \rho_k(v)$$

The algorithm may also be used to solve the inversion of noiseless observations $\hat{g}$ by choosing sequences ε_k, δ_k such that $\varepsilon_k/\delta_k \to 0$ and calculating for each ε_k, δ_k an approximation $G_{N(k)}$ by the above algorithm. Then in the limit $G_{N(k)}$ converges to the true G.

The construction of an algorithm for the inversion of Equation 2.6 is now complete. The salient features are: use of the singular value decomposition of $P_c \mathcal{F} P_c$, prior estimation of the number of degrees of freedom $N(\varepsilon,\delta, c)$ present at given noise levels, and construction of a Galerkin approximation from a predetermined set of functions of a simple closed form that approximate the first $N(\varepsilon,\delta, c)$ eigenfunctions.

A final note: g can not be approximated directly from the subspace spanned by $\{\mathcal{F} P_v \rho_k\}_{k=1}^{K(N)}$, as is often done with

$$\mathcal{F} P_c \rho_k(v) = \frac{\sin(v + k\pi/c)}{(v + k\pi/c)} \tag{3.27}$$

The untrustworthy components of the basis must first be filtered out by singular value decomposition before the extrapolation is done.

4. ITERATIVE SOLUTIONS AND ASSOCIATED THEORY

In contrast to the direct algorithm of the previous section we now consider iterative algorithms and integrate them into the body of theory already built up. We start with the original iterative algorithm, that of Gerschberg and Saxton [1]; a two-step iteration that updates the approximation in both physical and frequency space.

The formal statement of the algorithm applied to the model problem is:

a. Initial conditions

$$g_0 = G_0 = 0 \tag{4.1}$$

b. Update

$$g_{n+1} = P_{\bar{C}} \mathcal{F} P_C G_n + \hat{g} \tag{4.2}$$

$$G_{n+1} = P_C \mathcal{F}^{-1} g_{n+1} \tag{4.3}$$

where $\bar{C}$ is the set $\mathbb{R} - C$. An alternative, but more conventional, form is easily obtained. Concatenating the iterations above yields

$$G_{n+1} = P_C \mathcal{F}^{-1} (P_{\bar{C}} \mathcal{F} P_C G_n + \hat{g}) = (I - P_C \mathcal{F}^{-1} P_{\bar{C}} \mathcal{F} P_C) G_n + P_C \mathcal{F}^{-1} P_{\bar{C}} \hat{g} \tag{4.4}$$

upon noting that

$$P_C G_n = G_n \quad , \quad P_C \hat{g} = \hat{g} \tag{4.5}$$

$$P_C + P_{\bar{C}} = I \quad \text{(the identity operator)} \tag{4.6}$$

Equation 4.4 has a computational advantage over Equations 4.2, 4.3 because it requires evaluation of functions in the bounded set C, instead of the infinite domain $\bar{C}$, during the iteration.

Equation 4.4 is recognizable as the simplest form of iterative solution to the normal equations associated with Equation 2.6, that is

$$P_c \mathcal{F}^{-1} P_c \mathcal{F} P_c G = P_c \mathcal{F}^{-1} P_c \hat{g} \tag{4.7}$$

The normal equations are not usually formed for ill conditioned systems since they have even poorer conditioning; if $P_c \mathcal{F} P_c$ has singular values σ_n, then $P_c \mathcal{F}^{-1} P_c \mathcal{F} P_c$ has singular values σ_n^2. However since the singular values display the step like behavior, quoted in Equation 2.23, for almost all significantly non zero singular values $\sigma_n \approx \sigma_n^2 \approx 1$. Consequently little is lost by consideration of Equation 4.7.

In the presence of noise, the iteration based on the update Equation 4.4 is usually modified so that either it stops when

$$\| G_n - G_{n+1} \| \leq \delta \tag{4.8}$$

where δ depends on the noise level; or the damped update

$$G_{n+1} = (1 - \lambda)\, G_n - P_c \mathcal{F}^{-1} P_c \mathcal{F} P_c G_n + P_c \mathcal{F}^{-1} P_c \hat{g} \tag{4.9}$$

is used, where $\lambda > 0$ depends on the noise level. The update, Equation 4.9, if iterated to completion corresponds to a choice of filter

$$g(\sigma) = \frac{\sigma}{\sigma^2 + \lambda} \tag{4.10}$$

in Equation 2.20. However iteration to convergence is impossible in practice, and as yet there is no good theory linking termination after N steps with accuracy of approximation or noise levels. Therefore the clear insights singular value decomposition gave in modifying direct algorithms to take account of noise have no analog for iterative algorithms.

There is a large literature on alternative algorithms for iterative solutions of Equations 2.6 and 4.7, eg. [22]. The most attractive of these is the conjugate gradient method. This algorithm uses the residuals

$$r_n = P_c \mathscr{F}^{-1} P_c \hat{g} - P_c \mathscr{F}^{-1} P_c \mathscr{F} P_c G_n \tag{4.11}$$

in the following iteration scheme

a. Initial conditions

$$p_0 = G_0 = 0 \tag{4.12}$$

$$r_0 = P_c \mathscr{F}^{-1} P_c \hat{g} \tag{4.13}$$

b. Update

$$G_{n+1} = G_n + \alpha_n p_n \tag{4.14}$$

$$p_{n+1} = r_{n+1} + \beta_n p_n \tag{4.15}$$

where

$$\alpha_n \equiv \frac{\|r_n\|^2}{\|P_c \mathscr{F} P_c p_n\|^2} \quad , \quad \beta_n \equiv \frac{\|r_{n+1}\|^2}{\|r_n\|^2} \tag{4.16}$$

The conjugate gradient method stands midway between direct and iterative methods since it can be shown that the iterates G_N are also

Galerkin approximations for Equation 4.7 formed by taking $K = L = N$, $S_K = S_L = \text{span } \{r_k\}_{k=1}^{N}$. Although these approximations satisfy many of the criteria of section 3 (the r_k are orthogonal, $G_N \to G$ if G exists), an *a priori* estimate of amount of computation to achieve a desired accuracy cannot be made, nor can the iteration be modified in the presence of noise as is done for the direct methods.

If the iterative algorithms are run on a digital computer, a basis is implicitly used to represent the iterates. The only reason for not directly using this basis to form Galerkin approximations is cost. With present day computers, direct solutions are faster than iterative solutions to linear systems for all but the largest scale problems. This fact, combined with the theoretical results on the effective finite dimension of the model problem and a need for careful treatment of noise lead us to the conclusion that iterative algorithms are *not* the method of choice for this problem.

5. ILLUSTRATIVE NUMBERICAL CALCULATIONS

As an illustration of the results obtainable from the previous analysis, we present numerical solutions to two particular problems. The *first* is that of discriminating between possible bases for representation of g , G . The *second* is reconstruction of two particular $G(\omega)$ of interest in diffraction optics from noisy observations $\hat{g}(v)$ using the algorithm of Section 3.

Three commonly used bases for reconstruction are:

$$\rho_k^1(v) = \alpha_k P_k(v/c) \quad , \quad k = 0,1,\ldots,N \tag{5.1}$$

$$\begin{aligned} \rho_k^2(v) &= \beta_k \cos(k\pi v/2c) \quad , \quad k = \text{even} \\ &\qquad\qquad\qquad\qquad\quad k = 0,1,\ldots,N \\ &= \beta_k \sin((k+1)\pi v/2c) \quad , \quad k = \text{odd} \end{aligned} \tag{5.2}$$

$$\begin{aligned} \rho_k^3(v) &= \gamma_k \quad , \quad v \in \left[\frac{2ck - (N+1)c}{(N+1)} \quad , \quad \frac{2c(k+1) - (N+1)c}{(N+1)}\right], \\ &\qquad\qquad\qquad\qquad\quad k = 0,1,\ldots,N \\ &= 0 \quad , \quad \text{elsewhere} \end{aligned} \tag{5.3}$$

Here $P_k(x)$ is the kth Legendre polynomial, and α_k , ρ_k, γ_k are constants chosen so that the various ρ_k are orthonormal. All three bases may be seen as the initial segment of a complete basis for $L^2[-c,c]$. Van Buren [23] used ρ_k^1 to give accurate representations of the prolate spherioidal wave functions ϕ_k, the ρ_k^2 correspond to the sinc function basis of Equation 3.27

used widely in optics [24,25], and the ρ_k^3 are the familiar piecewise constant functions of numerical analysis (often used implicitly in algorithms based on sampled point values of $\hat{g}(v)$).

If a basis $\{\rho_k\}$ is used to generate approximate reconstructions via the algorithm of section 3, condition 1.c provides a natural merit function for the basis.

1. For fixed ε and c, the merit of the basis is the size of the least integer $K \equiv K(\varepsilon,c)$ such that

$$\| R_{KK} - P_c \mathcal{F} P_c \| < \varepsilon \tag{5.4}$$

However, as estimation of the norm in Equation 5.4 is too expensive for repeated calculation, we choose a slightly different merit function.

2. For fixed ε, c the merit of the basis is the size of the least integer $L \equiv L(\varepsilon,c)$ such that

$$|\sigma_n - s_n| < \varepsilon \quad , \quad \text{for all } n \tag{5.5}$$

where s_n are the singular values of R_{LL}.

Since the ρ_k are orthonormal, the s_n are also the singular values of the matrix $\tilde{D}$. This, together with knowledge of σ_n, allows calculation of $L(\varepsilon,c)$ for varying ρ_k, ε and c.

To rank the above bases, their $L(\varepsilon,c)$ were calculated for $c = .5, 1.0, 1.5, 2.0$ and $\varepsilon = .002, .02$. The results appear in Tables 1 and 2. $N(\varepsilon,c)$ denotes the number of σ_n greater than ε ; $L^i(\varepsilon,c)$, the $L(\varepsilon,c)$ of basis $\{\rho_k^i\}$. Because it is overly costly to consider every value of N , the table entries give an interval containing N rather than the exact result. The singular value decompositions of the matrices D were evaluated using the subroutine LSDVF from the IMSL library.

The tables indicate that $\{\rho_k^1\}$ has by far the highest merit rating. However, the complexity of programming and calculating the Legendre polynomials and their finite Fourier transforms the spherical Bessel functions $j_n(y)$, [26]

$$\int_{-1}^{+1} P_k(x)\, e^{iyx}\, dx = 2i^k\, j_k(y) \tag{5.6}$$

affects this rating. We feel their use is worthwhile only for large c or high accuracy computations. The ρ_k^2 are, suprisingly, the worst of the basis functions; it requires a very large number of such functions to approximate closely the higher order eigenfunctions ϕ_{2k+1}. We believe this is due to $\sin[(k+1)\pi v/2c]$ vanishing at the endpoints $v = \pm c$; however, $\phi_{2k+1}(\pm c) \neq 0$ consequently the ϕ_{2k+1}^2 produce poor approximations in this region. The same problem appears to a lesser degree in approximating even eigenfunctions ϕ_{2k}; at the endpoints the derivatives of $\cos(k\pi v/c)$ vanish whereas $d\phi_{2k}(\pm c)/dv \neq 0$. The low rating of the ρ_k^2 marks them as a poor choice for reconstruction and imply the sinc functions should not be used for extrapolation.

The basis of choice appears to be $\{\rho_k^3\}$, combining adequate approximation power with ease of computation. The basis also has natural extensions

to other problems. For greater accuracy, higher order splines may be preferable in use to $\{\rho_k^1\}$; for reconstruction problems in higher dimensions over arbitrarily shaped support sets A, B the basis extends to give the usual finite element type approximations.

These results on Galerkin approximations serve as a guide to construction of good discretizations based on pointwise quadrature rules. As an illustration we present some results on a discretization of Equation (2.5) on Gaussian quadrature. Let $\{p_i\}_{i=1}^N$ be the abscissae of the N point Gaussian quadrature scheme on $[-c,c]$ with associated weights $\{\omega_i\}_{i=1}^N$. Then Equation (2.5) may be approximated by

$$\hat{g} = \hat{F}\hat{W}\hat{G} \tag{5.7}$$

where $\hat{F}$ is an $N \times N$ matrix with entries

$$\hat{F}_{k\ell} = e^{2\pi i p_k p_\ell}, \tag{5.8}$$

W a diagonal matrix with entries $W_{kk} = \omega_k$ and $\hat{g}$ and $\hat{G}$ are vectors whose k-th entries are (hopefully) good approximations to $g(p_k)$ and $G(p_k)$.

However in this particular discretization the Euclidean norms of the vectors $\hat{g}$ and $\hat{G}$ bear little relation to the L^2 norms of $g(v)$ and $G(\omega)$ and calculations show that the singular values of $\hat{F}\hat{W}$ are not good approximations of those of $P_c \mathcal{F} P_c$. Therefore we replace Equation (5.7) by the system

$$(\hat{W}^{1/2}\hat{g}) = (\hat{W}^{1/2}\hat{F}\hat{W}^{1/2})(\hat{W}^{1/2}\hat{G}) \tag{5.9}$$

Now the Euclidean norms of $\hat{g}^1 \equiv \hat{W}^{1/2}\hat{g}$ and $\hat{G}^1 \equiv \hat{W}^{1/2}\hat{G}$ should coincide with

the L^2 norms of $g(v)$ and $G(\omega)$ and the singular values of $\hat{F}^1 \equiv (\hat{W}^{1/2}\hat{F}\hat{W}^{1/2})$ be close to those of $P_c \mathcal{F} P_c$.

We have calculated the $L(\varepsilon,c)$ of this approximation for the same values of ε and c used above, the results appear in Table 3. It appears that Gaussian quadrature is on a par with Galerkin approximations by Legendre polynomials as an efficient discretization; this is only to be expected given the role Legendre polynomials play in Gaussian quadrature. However in general Galerkin approximations seem preferable since they must be used anyway to choose among the many possible discretizations associated with a given quadrature rule, and to perform interpolation within the interval or extrapolation beyond it.

To test the actual inversion of noisy data, the algorithm of section 3, with $\{\rho_k^2\}$ as basis functions, was used to invert the following data corrupted with random noise:

$$g_1(v) = \frac{\sin 2\pi v}{\pi v} \quad , \quad |v| \leq 1 \tag{5.10}$$

$$g_2(v) = 2\left(\frac{\sin \pi v}{\pi v}\right)^2 \quad , \quad |v| \leq 1 \tag{5.11}$$

These functions can be interpreted as the point spread functions (≡ diffraction images) caused by an aberration-free slit aperture in coherent light and incoherent light respectively. The true solutions to the inversion problem are:

$$G_1(\omega) = 1 \quad , \quad |\omega| \leq 1 \tag{5.12}$$
$$= 0 \quad , \quad |\omega| > 1$$

$$G_2(\omega) = 2(1 - |\omega|) \quad , \quad |\omega| \leq 1$$
$$= 0 \qquad\qquad\quad |\omega| > 1 \tag{5.13}$$

The G's are the corresponding optical transfer functions in the diffraction imagery interpretation. In both cases $c = 1$. The calculated inversions were compared to the true $G_i(\omega)$ on $|\omega| \leq 1$, and the approximate extrapolations of Equation (3.26) were compared to the true $g_i(v)$ on $|v| \leq 4$.

We first applied the algorithm to noiseless data, choosing $\delta = 1$, $\varepsilon = 10^{-2}$ and $K(N) = 40$ for $\hat{g}_1$, $K(N) = 80$ for $\hat{g}_2$ (in both cases $\|R_{K(N)K(N)} - P_c \mathcal{F} P_c\| < \varepsilon$). The resulting approximations to $G_i(\omega)$ coincides with the true $G_i(\omega)$ to well within graphical accuracy. The approximate extrapolation to $g_i(v)$ also coincided to the true $g_1(v)$ to within graphical accuracy. The approximate extrapolation to $g_2(v)$ is plotted in Figure 1, the solid line is the true $g_2(v)$.

To simulate the presence of noise in the data, whenever a value of $g(v)$ was required in the numerical integrations used to calculate the components $c_k = (g, \rho_k^2)$, a random perturbation was added. That is

$$g_\varepsilon(v) = g(v) + \varepsilon\mu \max_{v \in c} |\hat{g}(v)| \tag{5.14}$$

was used in place of $g(v)$. The constant ε denotes the noise level, μ is a random variable uniformly distributed over $[-1,1]$. For this artificial noise, the error level ε of the algorithm was taken to be the ε in Equation (2.8). Since we had no prior expectations on the accuracy of the inversions, the error tolerance δ of the algorithm was taken to be unity.

To test the *robustness* of the algorithm, reconstruction of data with different random perturbations of different noise levels were made. The same K(N) as in the noiseless case were used. Table 4 lists four reconstructions of $G_1(\omega)$ from data with 1% (i.e., $\varepsilon = .01$) noise level; there is approximately a 5% maximum deviation from the true value. Figure 2 shows four reconstructions of $G_1(\omega)$ from data with a 3% noise level; there is now a 25% maximum deviation from the true value. Reconstructions from data with 5% noise levels were still recognizable; but at 10% noise levels the error (measured in the energy norm) in the reconstruction exceeded 50%. Figures 3 and 4 show (see open circles) the extrapolation of two sample reconstructions from Figure 1. For comparison the true $g_1(v)$ is also shown as the solid line. Because G_{i-} and g_i are even functions of their arguments, the graphs have been plotted only for the negative values of the respective arguments.

Figure 5 and 6 show four realizations of $G_2(\omega)$ from data with 3% noise level. Extrapolations of three of these reconstructions appear in Figures 7-9, with the true $g_2(v)$ shown as the solid line. Although the reconstructions of $G_2(\omega)$ appear to be more accurate than those of $G_1(\omega)$, the extrapolation to $g_2(v)$ is less accurate than those to $g_1(v)$. This is due to the discontinuity in slope of $G_2(\omega)$ at the origin. Since the approximations to $G_2(\omega)$ are sums of small numbers of smooth eigenfunctions, they do not approximate $G_2(0)$ well as the graphs indicate. This low frequency error in the ω domain is then transformed into high frequency errors in the v domain, resulting in errors in the extrapolation.

The analysis used in defining the "trustworthy" component of G, and a trustworthy Galerkin approximation, is *worst case* in nature. It is therefore reasonable to supply the algorithm with smaller ε or larger δ than *a priori* noise levels suggest (as was done here) or use a different filter in Equation (3.25). However this cannot be made precise without some statistical statements on expected errors.

ACKNOWLEDGMENTS

We are indebted to Professor D.G.M. Anderson for helpful criticisms of this work. Barakat was supported in part by RADC, Newsam was supported in part by NSF.

REFERENCES

1. W.O. Saxton, Computer Techniques for Image Processing in Electron Microscopy (Academic Press, New York, 1978) Chpts. 5 and 6.

2. A. Papoulis, "A new algorithm in spectral analysis and bandlimited extrapolation." IEEE Trans. Circuits Syst. CAS22, 735-742 (1975).

3. D.C. Youla, "Generalized image restoration by the method of alternating orthogonal projections." IEEE Trans. Circuits Syst., CAS25, 695-702 (1978).

4. J.A. Cadzow, "An extrapolation procedure for bandlimited signals." IEEE Trans. Acoust. Speech Signal Process., ASSP27, 4-12 (1979).

5. M.S. Sabri and W. Steenart, "An approach to bandlimited extrapolation: the extrapolation matrix." IEEE Trans. Circuits Syst., CAS25, 74-78 (1978).

6. R.E. Burge, M.A. Fiddy, A.H. Greenway, G. Ross, "The phase problem." Proc. Roy. Soc., A350, 191-212 (1976).

7. J.M. Oretega and W.C. Rheinboldt, Iterative Solution of Nonlinear Equations in Several Variables (Academic Press, New York, 1970) pp. 494-500.

8. R. Barakat and G. Newsam, "Numerically stable iterative method for the inversion of wavefront aberrations from measured point spread function data." J. Opt. Soc. Amer., 70, 1255-1263 (1980).

9. J.R. Fienup, "Space-object imaging through atmosphere," Opt. Eng., 18, 529-534.

10. A.N. Tikhonov and V.Y. Arsenin, Solutions of Ill-Posed Problems (Halsted, New York, 1977) pp. 7-9.

11. R.P. Boas, Entire Functions (Academic Press, New York, 1954) p. 103.

12. E.J. Akutowicz, "On the determination of the phase of a Fourier integral, I." Trans. Amer. Math. Soc., 83, 179-192 (1956).

13. E.J. Akutowicz, "On the determination of the phase of a Fourier integral, II." Proc. Amer. Math. Soc., 8, 234-238 (1957).

14. J.N. Chapman, "The application of iterative techniques to the investigation of strong phase objects in the electron microscope." Phil. Mag., 32, 527-552 (1975).

15. M. Bendinelli, A. Consortini, L. Ronchi, and B. Frieden, "Degrees of freedom, and eigenfunctions, for the noisy image." J. Opt. Soc. Amer., 64, 1498-1502 (1974).

16. F. Gori and G. Guattari, "Shannon number and degrees of freedom of an image." Opt. Comm., 7, 163-165 (1973).

17. C.T.H. Baker, The Numerical Treatment of Integral Equations (Clarendon Press, Oxford, 1977).

18. G.W. Stewart, Introduction to Matrix Computations (Academic Press, New York 1973) pp. 317-326.

19. D. Slepian and H. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty - I." Bell Syst. Tech. J., 40, 43-63 (1961).

20. H. Landau and H. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty - II." Bell Syst. Tech. J., 40, 65-84 (1961).

21. H. Landau and H. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty - III." Bell Syst. Tech. J., 41, 1295-1336 (1962).

22. W.M. Patterson, III, Iterative Methods for the Solution of a Linear Operator Equation in Hilbert Space - A Survey (Springer-Verlag, Berlin, 1974).

23. A.L. Van Buren, A Fortran Computer Program for Calculating the Linear Prolate Functions, NRL Report 7994 (Naval Research Laboratory, Washington, D.C., 1976).

24. R. Barakat, "Application of the sampling theorem to optical diffraction theory," J. Opt. Soc. Amer., 55, 920-927 (1964).

25. A. Jerri, "The Shannon sampling theorem - its various extensions and applications: a tutorial review." Proc. IEEE, 65, 1565-1596 (1977).

26. A. Erdelyi, et al, Tables of Integral Transforms, Vol. 1 (McGraw-Hill, New York, 1954) p. 122.

TABLE 1: Distribution of singular values, and required basis size for desired accuracy when $\varepsilon = .02$

c	$N(\varepsilon,c)$	$L^1(\varepsilon,c)$	$L^2(\varepsilon,c)$	$L^3(\varepsilon,c)$
.5	3	3	20 - 30	3 - 10
1.0	7	7 - 9	40 - 50	10 - 20
1.5	13	13 - 15	>80	20 - 40
2.0	20	23 - 25	>80	40 - 80

TABLE 2: Distribution of singular values, and required basis size for desired accuracy when $\varepsilon = .002$

c	$N(\varepsilon,c)$	$L^1_\cdot(\varepsilon,c)$	$L^2(\varepsilon,c)$	$L^3(\varepsilon,c)$
.5	4	5 - 7	>80	10 - 20
1.0	9	9 - 11	>80	40 - 80
1.5	15	19 - 23	>80	>80
2.0	22	31 - 35	>80	>80

TABLE 3. Distribution of $L(\varepsilon,c)$ for approximations based on Gaussian quadrature.

c	L(.02,c)	L(.002,c)
.5	4-8	4-8
1.0	8-12	8-12
1.5	14-18	18-22
2.0	30-34	30-34

TABLE 4: Four Sample Realizations of Reconstructions with 1% Noise.

ω	$G(\omega)$	$G(\omega)$	$G(\omega)$	$G(\omega)$
-1.0000	.983	.973	1.099	.907
- .9474	1.009	1.010	1.043	1.005
- .8947	1.013	1.021	.995	1.046
- .8421	1.004	1.017	.962	1.050
- .7895	.991	1.007	.944	1.035
- .7368	.980	.997	.942	1.013
- .6842	.974	.990	.952	.993
- .6316	.975	.989	.971	.980
- .5789	.980	.993	.995	.976
- .5263	.988	1.000	1.017	.979
- .4737	.999	1.008	1.035	.988
- .4211	1.007	1.015	1.046	.998
- .3684	1.013	1.019	1.048	1.007
- .3158	1.014	1.019	1.042	1.012
- .2632	1.011	1.015	1.028	1.014
- .2105	1.005	1.008	1.008	1.012
- .1579	.997	.999	.987	1.007
- .1053	.988	.990	.968	1.001
- .0526	.982	.983	.953	.996
0	.978	.979	.945	.993

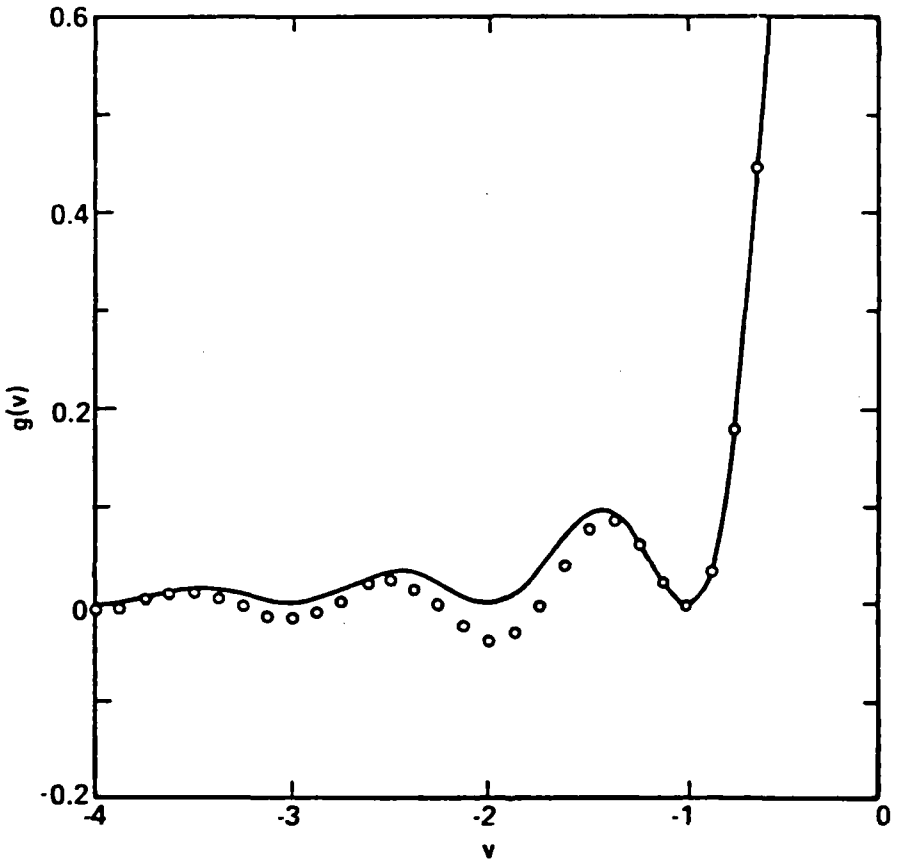

Figure 1: Extrapolation of $g_2(v)$, see open circles, corresponding to a noiseless situation.

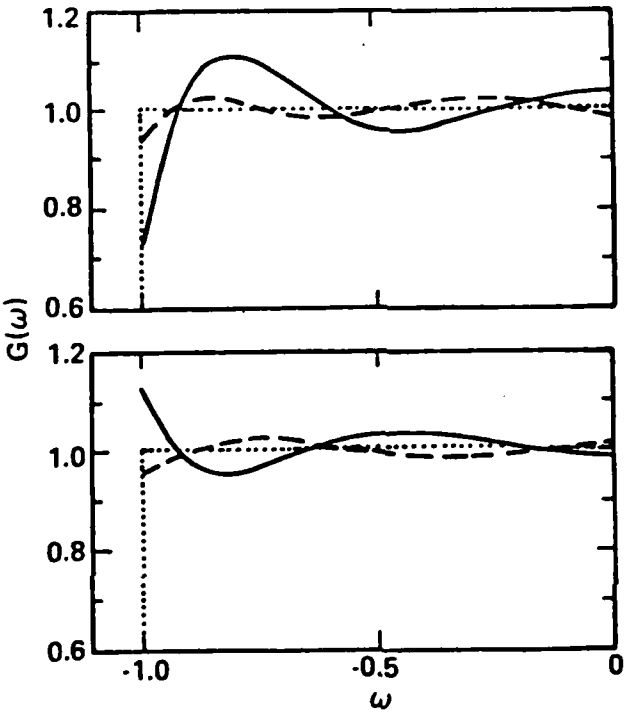

Figure 2: Four sample realizations of the reconstruction of $G_1(\omega)$ in the presence of 3% noise in $\hat{g}(v)$.

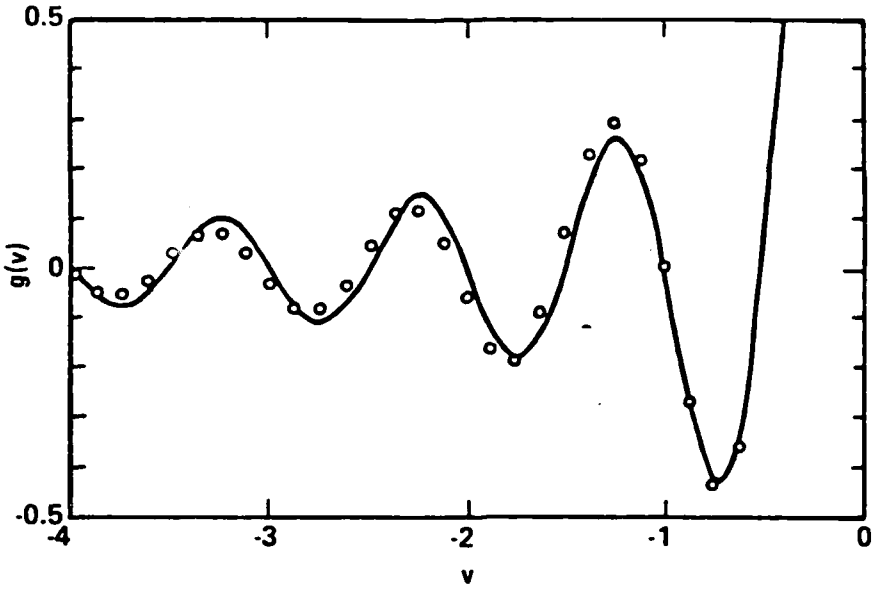

Figure 3: Sample realization extrapolation of $g_1(v)$ corresponding to solid line in upper half of Figure 2.

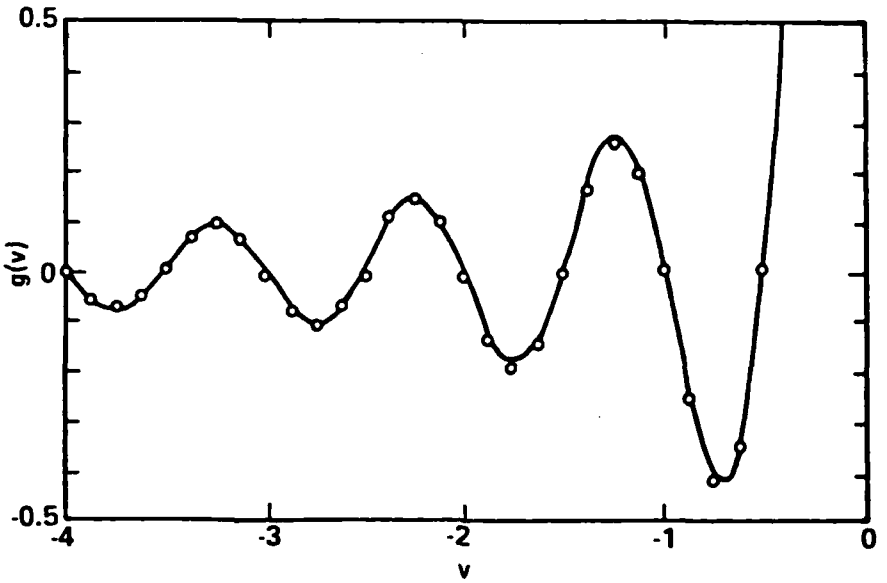

Figure 4: Sample realization extrapolation of $g_1(v)$ corresponding to dashed line in upper half of Figure 2.

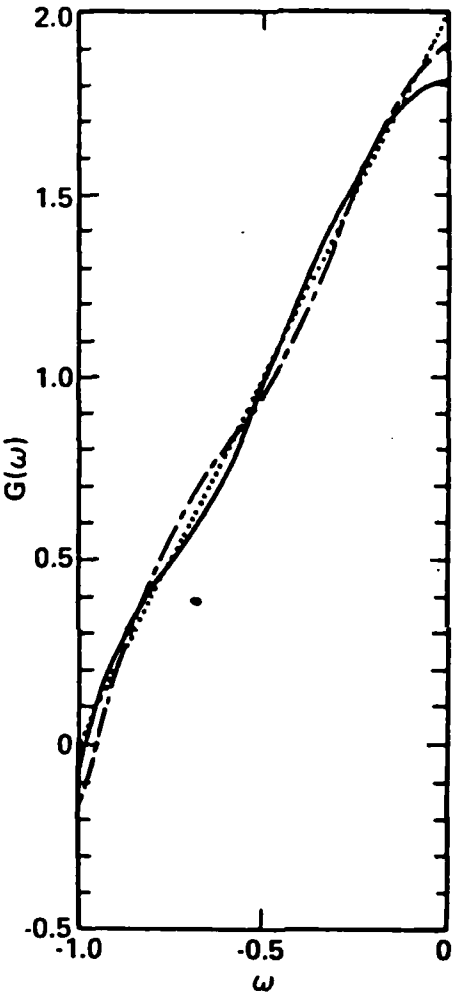

Figure 5: Two sample realizations (— • — and ——) of the reconstruction of $G_2(\omega)$ in the presence of 3% noise in $\hat{g}(v)$.

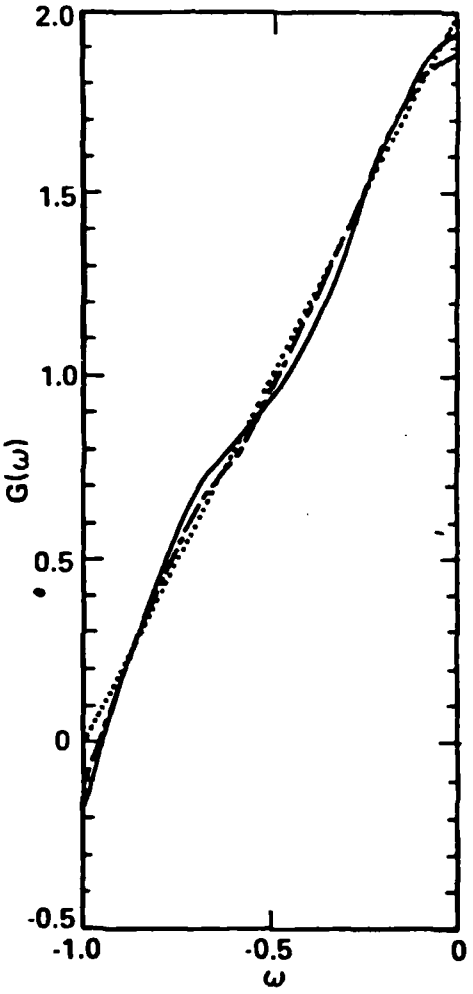

Figure 6: Two sample realizations (— • — and ——) of the reconstruction of $G_2(\omega)$ in the presence of 3% noise in $\hat{g}(v)$.

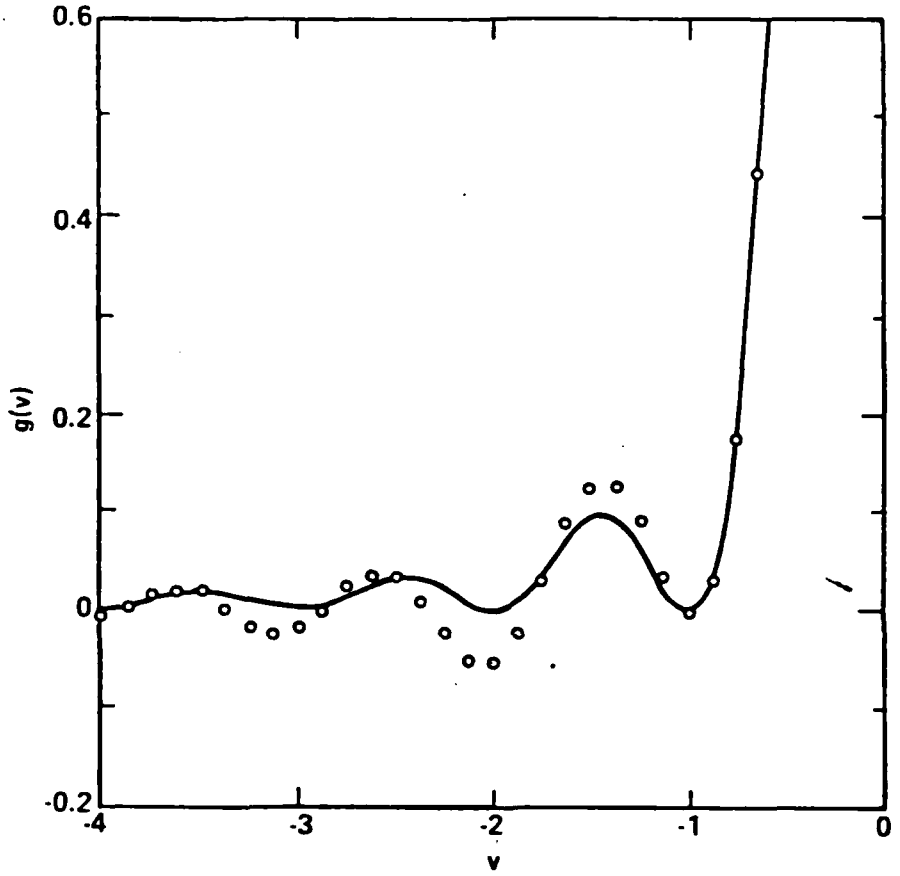

Figure 7: Sample realization extrapolation of $g_2(v)$ corresponding to —— line in Figure 5.

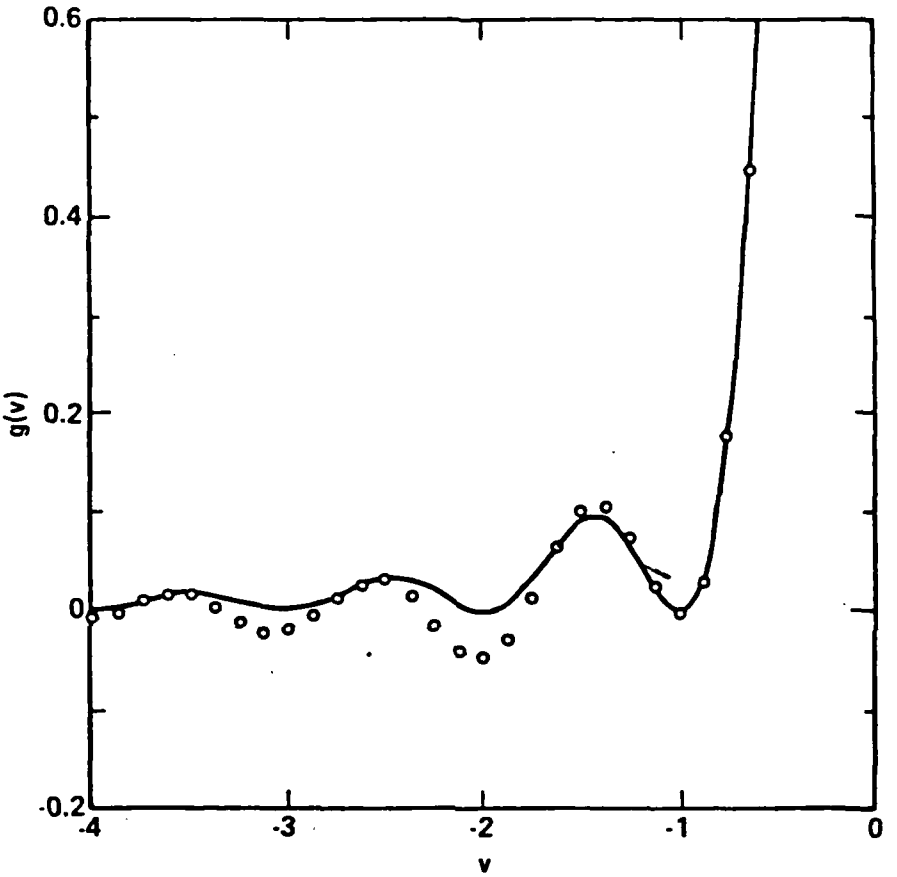

Figure 8: Sample realization extrapolation of $g_2(v)$ corresponding to —•— line in Figure 5.

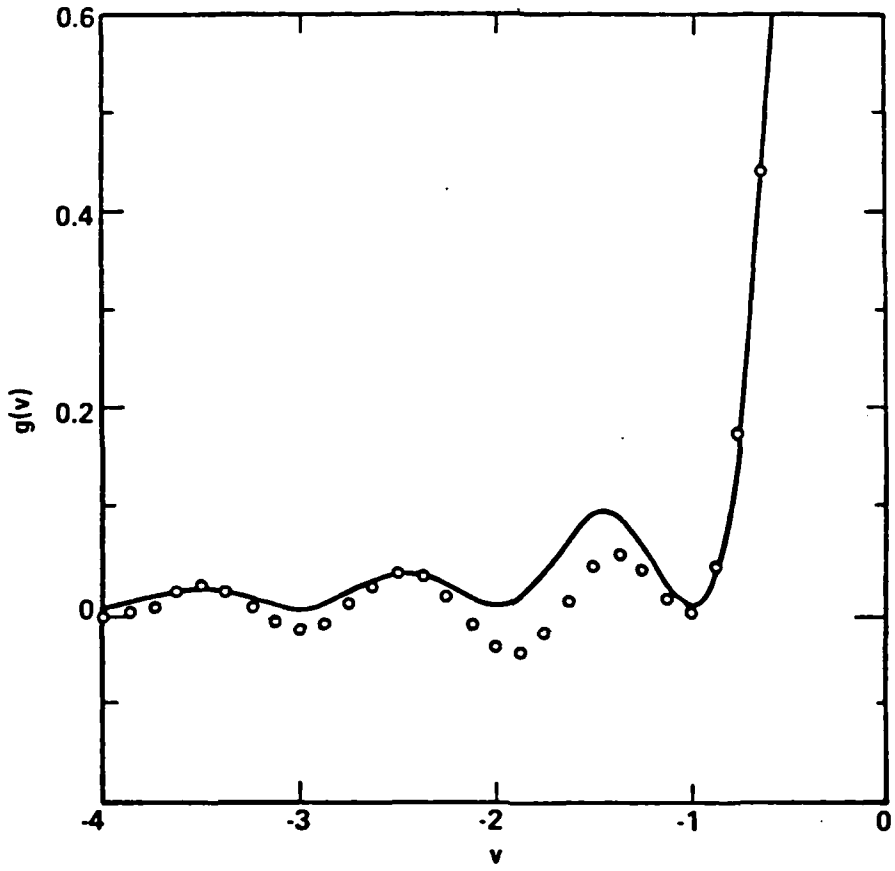

Figure 9: Sample realization extrapolation of $g_2(v)$ corresponding to —•— line in Figure 6.

ALGORITHMS FOR RECONSTRUCTION OF PARTIALLY KNOWN, BANDLIMITED FOURIER TRANSFORM PAIRS FROM NOISY DATA: II THE NONLINEAR PROBLEM OF PHASE RETRIEVAL

ABSTRACT

Phase retrieval problems are ill-posed, however previous analysis has focused upon global ill conditioning due to the existence of multiple exact solutions. In this paper we consider the effects of local ill conditioning due to the presence of large infinite dimensional neighborhoods of any exact solution where members are all possible solutions provided there is any uncertainty in the data. The form of such neighborhoods can be characterized by viewing phase retrieval as a nonlinear extension of the linear problem of inversion of the finite Fourier transform considered in the previous companion to the present work. In particular, we are able to estimate the essential dimension of phase retrieval problems, i.e. the number of parameters in a solution representation that can be determined accurately given specified error levels in the data. Based on these results a modification of the widely used Gerschberg-Saxton algorithm is proposed, analyzed, and then used as a basis for development of more sophisticated algorithms. Numerical results are presented on the performance of these algorithms on one-dimensional problems. The results indicate that although the algorithms may be tuned to overcome local ill conditioning, good solutions in one dimension are still difficult to find numerically because of global ill conditioning. However, the material in the Appendix indicates that in higher dimensions global ill conditioning is considerably reduced so that the algorithms should be effective in such problems.

1. INTRODUCTION

In our companion paper [1], we introduced the problem of numerical reconstruction of a partially known, bandlimited Fourier transform pair g, G and considered in detail the model linear problem

> Given values over a finite interval $[a_1,a_2]$ of a function $g(v)$ known to be the Fourier transform of a function $G(\omega)$ with support contained in the finite interval $[b_1,b_2]$, find $G(\omega)$.

The paper focused on the ill posed nature of this problem and its effects on numerical solution, an algorithm (filtered singular-value decomposition) was presented that could be specially tailored to cope with these effects. In this paper we consider the associated model *nonlinear* problem of *phase retrieval*.

The model problem in phase retrieval is

> Given value $m(v)$ over a finite interval $[a_1,a_2]$ of the modulus of a function $g(v)$ known to be the Fourier transform of a function $G(\omega)$ with support contained in the finite interval $[b_1,b_2]$ and the knowledge that G is a member of some set B, find $g(v)$ and $G(\omega)$.

The set B represents available prior knowledge on G. In this paper we study three particular cases: no prior knowledge, G is nonnegative, and $|G(\omega)| \equiv n(\omega)$ is known over $[b_1,b_2]$. Again our focus is on the ill posed nature of phase retrieval and the resulting implications for numerical reconstructions. In particular we generalize the iterated projection algorithm of Gerschberg and Saxton [2] in such a fashion so that it may be identified

as a first order, gradient type optimization algorithm. We then propose second order, Newton type extensions of the algorithm and show that the linear problem arising at each step has behaviour dominated by the behaviour of the linear problem [1].

The structure of the paper is as follows. Section 2 sets out the notation employed in the paper and lists the particular model problems used in illustrative numerical calculations. Section 3 contains a discussion of the ill conditioning of phase retrieval, focusing on the two causes: possible multiple solutions and ill conditioning of the underlying linear operator (the finite Fourier transform). In the Appendix we show that although in one dimension multiplicity of solutions presents a serious problem, in two or more dimensions its occurrence is pathologically rare. Section 4 is devoted to a generalization of the Gerschberg-Saxton algorithm through restatement of the phase retrieval problem as one of finding the closest point to several, possibly disjoint, sets. Two particular algorithms are developed that use iterated projections; their convergence properties are investigated and their identification with the well known steepest descent algorithm is established. In Section 5 a number of second order algorithms are developed from those of Section 4. Each algorithm requires the solution of an ill posed linear problem at each iteration; we propose that this solution be found by filtered singular value decomposition. In order to save on the cost of an expensive decomposition calculation at each step, we indicate how a partial filtering can be obtained using a block decomposition of the linear system and a precalculated singular value decomposition of the

finite Fourier transform (fFT) that is constant across iterations. Finally, Section 6 contains numerical results showing the application of these algorithms to each of the model phase retrieval problem. Tables showing the relative performance of the various algorithms and graphs of the approximate solutions generated are presented and discussed.

The results indicate that in one dimension the multiplicity of solutions induces poor global conditioning so that the significant improvement of second order methods over first order methods is restricted to small local neighborhoods of the solution. Therefore the simplest algorithms appear to be the most cost effective, however we feel that in higher dimensions, with the increased likelihood of unique solutions, that higher order methods will come into their own.

2. SOME MODEL PROBLEMS IN PHASE RETRIEVAL

The analysis of this paper is set in the function space $L^2(\mathbb{R})$ with the standard inner product $(\cdot,\cdot)$ and norm $\|\cdot\|$. The Fourier transform $\mathcal{F}: L^2(\mathbb{R}) \to L^2(R)$ is here defined to be

$$g(v) = (\mathcal{F}g)(\omega) = \int_{-\infty}^{\infty} e^{2\pi i v\omega} G(\omega)\, d\omega \quad . \tag{2.1}$$

Subsets of $L^2(\mathbb{R})$ will be denoted by A_i or B_j, and the projection operator $P_A: L^2(\mathbb{R}) \to A$ onto a *particular* set A is defined by

$$P_A f = g \iff \|f-g\| = \min_{h \in A} \|f-h\| \tag{2.2}$$

$P_A f$ is termed the projection of f on A, projections are assumed to exist and be unique.

The interval $[-1,1]$ will be denoted by I and $[-c,c]$ by cI; cI will also be used to denote the subset of $L^2(\mathbb{R})$ of all functions whose support is contained in $[-c,c]$. The associated projection operator is therefore

$$\begin{aligned} (P_{cI} f)(v) &= f(v) \ , \qquad \text{if} \quad v \in [-c,c] \\ &= 0 \quad , \qquad \text{if otherwise} \ . \end{aligned} \tag{2.3}$$

It will be abbreviated to P_c for typographic convenience.

Three model problems, variations of phase retrieval that occur in physical problems, will be used as numerical illustrations in the remaining sections. They are:

I. Given values $m(v)$ over the interval $[a_1,a_2]$ of the modulus of a function $g(v)$ known to be the Fourier transform of a function $G(\omega)$ whose support is contained in the interval $[b_1,b_2]$, find $g(v)$ and $G(\omega)$. Some relevant references are [2,3].

II. Given values $m(v)$ over the interval $[a_1,a_2]$ of the modulus of a function $g(v)$ known to be the Fourier transform of a real, nonnegative function $G(\omega)$ whose support is contained in the interval $[b_1,b_2]$, find $g(v)$ and $G(\omega)$. Some relevant references are [4.5].

III. Given values $m(v)$ over the interval $[a_1,a_2]$ of the modulus of a function $g(v)$ known to be the Fourier transform of a function $G(\omega)$ whose support is contained in the interval $[b_1,b_2]$ and whose modulus $n(\omega)$ is given over $[b_1,b_2]$, find $g(v)$ and $G(\omega)$. Some relevant references are [6,7].

These two moduli $m(v)$ and $n(\omega)$ may only be known to within some accuracy ε, that is *measurements* $\tilde{m}$ and $\tilde{n}$ are available such that

$$\|m-\tilde{m}\|, \qquad \|n-\tilde{n}\| \leq \varepsilon \ . \tag{2.4}$$

In this setting all problems and errors are invariant under translation and scaling; so, as in the previous paper [1], the sets $[a_1,a_2]$, $[b_1,b_2]$ can be transformed into cI where

$$c = \frac{1}{2}\,[(a_2 - a_1)(b_2 - b_1)]^{1/2} \ . \tag{2.5}$$

Upon defining the sets

$$A \equiv \{g: |(P_c g)(v)| = (P_c m)(v)\} \tag{2.6a}$$

$$B_1 \equiv cI \tag{2.6b}$$

$$B_2 \equiv cI \cap \{G: G \geq 0\}. \tag{2.6c}$$

$$B_3 \equiv cI \cap \{G: |(P_c G)(\omega)| = (P_c n)(\omega)\} \tag{2.6d}$$

the i-th problem $(i = 1,2,3)$ can be rewritten as:

$$\text{Find } g,G \text{ such that } g = \mathscr{F}G, \quad g \in A, \; G \in B_i \quad . \tag{2.7}$$

3. THE ILL-POSED NATURE OF PHASE RETRIEVAL

In the previous paper we indicated that the phase retrieval problem is, in general, ill-posed in that it fails to satisfy Hadamard's definition of a well-posed problem (the definition is repeated here for convenience).

Definition: A problem is well-posed if the solution

a) exists

b) is unique

c) depends continuously on the data.

If any of these conditions are violated it is ill-posed.

We now consider the ill-posed and ill conditioned nature of the phase retrieval problem in more detail with particular regard as to the consequences for numerical solutions. We wish to show that it is a failure of condition c that is the main source of difficulty.

To illustrate the point we briefly review the prototypical linear problem and its solution as described in the previous paper, the inversion of a compact linear operator by use of its filtered singular value decomposition (SVD). Let H be a separable Hilbert space and $\mathcal{K}:H \to H$ be a compact linear operator with an SVD$\{\phi_i,\sigma_i,\psi_i\}$. If g is an arbitrary member of H with the expansion

$$g = \sum_{i=0}^{\infty} a_i\psi_i \quad , \qquad \sum_{i=0}^{\infty} |a_i|^2 < \infty \tag{3.1}$$

then the equation $\mathcal{K}G = g$ has the formal solution

$$G = \sum_{i=0}^{\infty} \left(\frac{a_i}{\sigma_i}\right) \phi_i \quad . \tag{3.2}$$

The solution G is in H if and only if $\sum_{i=0}^{\infty} |a_i \sigma_i^{-1}|^2 < \infty$. Because the singular values σ_i tend to zero this is a stronger condition on the coefficients a_i than Eq. (3.1), so a solution does not exist for every right hand side g. Thus the problem is ill-posed with respect to existence of solutions. The solution is unique if and only if $\mathcal{K}$ has a non-trivial null space, which is true if and only if $\sigma_i > 0 \ \forall i$. The solution does not depend continuously on the data. If a small perturbation ε is made in a high frequency component a_i of g then a perturbation (ε/σ_i) is induced in G; this can be made as large as desired by increasing i. Thus for any given ε, δ and solution pair g, G there exists an infinite dimensional set of solutions $\tilde{g}$, $\tilde{G}$ such that

$$\mathcal{K}\tilde{G} = \tilde{g}, \quad \|g - \tilde{g}\| < \varepsilon, \quad \|G - \tilde{G}\| > \delta \quad . \tag{3.3}$$

It is violation of condition c that makes numerical inversion of compact operators so difficult. In problems arising from physical systems existence of a solution is a prerequisite for making the measurements; failure of the subsequent mathematical model to have a solution is usually due to measurement errors in g or $\mathcal{K}$. Thus non-existence is really a consequence of discontinuous dependence and unimportant in its own right. Likewise the existence of multiple *exact* solutions is not in itself a problem. In most cases the nullspace can be predetermined theoretically and the operator restricted to the complement of this space. If the restricted

operator had a bounded inverse, numerical inversion would present no further difficulties.

Solution by filtered SVD directly addresses the problem of discontinuous dependence on the data by identifying the infinite dimensional subspace S over which $\mathcal{K}$ produces distinct (but very small) variations; then restricting the approximate solutions to the complementary subspace $S^{\perp}$ over which $\mathcal{K}^{-1}$ is uniformly continuous. This decomposition, effected by the filter, corresponds to choosing the first N terms in a generalized Fourier expansion of the solution G; as such it has a wide variety of uses and interpretations.

In the nonlinear problem of phase retrieval the traditional investigations of ill conditioning and ill positioning have concentrated on the questions of existence and uniqueness [3]. Necessary and sufficient conditions that m and g must satisfy for existence of a solution G are given by the Paley-Wiener theorem. For one-dimensional retrieval problems the form and existence of multiple solutions is a consequence of the Hadamard factorization theorem (a result first derived by Akutowicz [8 , 9]. An extension of this theorem gives necessary conditions for existence of multiple solutions in two-dimensional problems. A detailed exposition appears in Appendix A.

Although questions of existence and uniqueness have some influence on the behavior of numerical algorithms, as with the linear problem discontinuous dependence on the data dominates. Existence or non-existence of solutions is again just a simple consequence of discontinuous dependence. If multiple solutions exist but are uniformly separated, their existence will influence

the initial iterations of any algorithm but the final behavior will be determined by the question of whether the problem is well posed in a neighborhood of each solution or not. As we shall show the phase problem is never well posed.

However existence of multiple solutions may be a stronger contributor to ill conditioning than indicated above. As shown in the first section of Appendix A, in the one-dimensional problem there may exist an infinite sequence of exact solutions G_N with a limit G_∞, corresponding to a sequence of finite products of Blaschke factors and the limiting infinite product. In this case the solutions are not uniformly separated and any algorithm will have difficulties in the neighborhood of G_∞. Fortunately, for the reasons outlined in Appendix A, in two-dimensional problems multiple solutions appear to be rare and the limiting behavior described above unlikely in the extreme, unless symmetry considerations reduce the problem to an essentially one-dimensional form.

Proof of discontinuous dependence on the data is difficult for an arbitrary nonlinear equation $\mathscr{L}(G) = g$. The most widely used criterion is that the Frechet derivative $D(G)$ of $\mathscr{L}$ at the point G be a compact linear operator. The phase retrieval problem may be formally stated as that of finding a solution pair to the equation

$$\mathscr{L}(G) \equiv |P_c \mathscr{F} P_c G| \equiv (\mathscr{L}_1 \circ P_c \mathscr{F} P_c)(G) = |P_c g| = m \tag{3.4}$$

where the nonlinear operator $\mathscr{L}_1$ is defined by

$$\mathscr{L}_1(g) \equiv |g| \tag{3.5}$$

so that $\mathscr{L}$ is the composition $(\circ)$ of $\mathscr{L}_1$ and the compact linear operator $P_c \mathscr{F} P_c$. The Fréchet derivative $D_1(g)$ of $\mathscr{L}_1(g)$ is the bounded linear operator defined by

$$D_1(g)h \equiv \mathrm{Re}(g^*.h)/|g| \quad . \tag{3.6}$$

The Fréchet derivative $D(G)$ of $\mathscr{L}(G)$ is the operator $D_1(G) \circ P_c \mathscr{F} P_c$; since this is the composition of a bounded operator with a compact operator it follows that $D(G)$ is compact and the phase retrieval ill-posed.

Since $D(G)$ varies with G the decomposition by filtering of the underlying space into a subspace $S(G)$, over which $D(G)$ is slowly varying, and its complement $S^{\perp}(G)$, is not a global decomposition. Instead $S(G)$ and $S^{\perp}(G)$ vary with G and serve as tangent planes in the definition of manifolds $\mathscr{M}$ and $\mathscr{M}^{\perp}$ over which $\mathscr{L}(G)$ is slowly varying or $\mathscr{L}^{-1}$ is uniformly continuous. However as $\mathscr{M}^{\perp}$ is not linear, generation of an approximate solution $\tilde{G} \in \mathscr{M}^{\perp}$ does not convey the information that is contained in generation of an approximate solution from a linear subspace $S^{\perp}$; and is also a much harder problem.

The usual approach taken to overcome this problem, and to deal with measurement noise, is to reduce the ill posed problem to a well posed problem by regularization; that is a parameter $\lambda > 0$ and functional $\Omega(G)$ are chosen and the regularized solution G_λ found by minimization of

$$\| \mathscr{L}(G) - y \| + \lambda \Omega(G) \tag{3.7}$$

is taken as an approximation to G. As stated the parameter λ and functional Ω do not appear to depend on measurement noise or the problem

form in any obvious way, but the following theorem due to Tikhonov [10] elucidates their relationship

Theorem 1: Under certain mild conditions on $\mathscr{L}$, Ω, G and g, then

$$G_\lambda \text{ minimizes } \|\mathscr{L}(G) - g\| + \lambda\Omega(G)$$

if and only if

a) there exists $\delta_\lambda > 0$ such that

$$G_\lambda \text{ minimizes } \|\mathscr{L}(G) - g\| \text{ subject to } \Omega(G) \leqslant \delta_\lambda$$

b) there exists $\varepsilon_\lambda > 0$ such that

$$G_\lambda \text{ minimizes } \Omega(G) \text{ subject to } \|\mathscr{L}(G) - g\| \leqslant \varepsilon_\lambda .$$

This approach is explicit or implicit in many of the algorithms presented in the literature [11].

Solution by regularization does not identify an approximating subspace, or even a submanifold, such as that found by filtering; settling rather for the (less informative) identification of sets $A_\varepsilon \equiv \{G: \|L(G) - g\| \leqslant \varepsilon\}$ and $A_\delta \equiv \{G: \Omega(G) \leqslant \delta\}$ that contain the solution G. However the phase problem does have a natural decomposition of the solution space into subspaces rather than submanifolds, encouraging solution by filtered SVD rather than by regularization. This follows from the form of D(G). Upon definition of the subspace

$$T \equiv \{h: \arg h = \arg g\} \tag{3.8}$$

it is easily shown that

$$\begin{aligned} D_1(g)h &= e^{-i(\arg g)}h && h \in T \\ &= 0 && h \in T^\perp \end{aligned} \tag{3.9}$$

Consequently $D_1(g)$ has a well defined null space $T^{\perp}$ and a bounded inverse on the complement T. Therefore the ill conditioning of D(G) is essentially due to the operator $P_c \mathcal{F} P_c$ studied in the previous paper. This suggests that the global decomposition of H into linear subspaces S and $S^{\perp}$ used there in inversion of $P_c \mathcal{F} P_c G = g$ by filtered SVD will also be appropriate in the nonlinear problem. This idea is further developed in Section 6.

4. ITERATIVE METHODS

The formulation of phase retrieval as the search for a common intersection point, given in Equation (2.7), suggests use of successive projection algorithms for its solution. The simple structure of the projection operators P_A and P_{B_i} for the sets A and B_i appearing in Eq. (2.7) indicates that this class of algorithms will be computationally efficient. Such algorithms are presented for the general problem in [12] and were first applied to phase retrieval by Gerschberg and Saxton [7,13]. However these algorithms were originally derived and analyzed under conditions such as existence of an intersection point, so they should not be applied directly to the phase problem due to the presence of ill conditioning. We therefore propose and discuss modifications more suited to the ill-posed nature of this problem.

4.1 *The Generic Algorithm*

Let B_i be a collection of M sets with associated projection operators P_i. Since ill conditioning implies that $\bigcap_{i=1}^{M} B_i$ can be empty, we attempt to find an x that is closest to all sets B_i and which reduces to a common intersection point *if one exists*. Therefore we seek to minimize

$$F(x) \equiv \sum_{i=1}^{M} \|x - P_i x\|^2 \tag{4.1}$$

The simplest iterative algorithms for minimization of $F(x)$ can be considered as pointwise approximation. For an arbitrary point x the points $b_i = P_i x$ are point approximations to sets B_i and induce an approximation

$$G(z,x) = \sum_{i=1}^{M} \|z - b_i\|^2 \tag{4.2}$$

to $F(x)$. The algorithm then constructs a sequence x_n, where x_{n+1} is determined from x_n after minimization of $G(z,x_n)$.

Lemma 1: $G(z,x)$ has a unique minimum at $y = M^{-1} \sum_{i=1}^{M} P_i x$ and if $y \neq z$ then $G(z + \lambda(y - z),x) < G(z,x)$ $\forall \lambda \in (0,2)$.

Proof:

$$\begin{aligned} G(z,x) &= \sum_{i=1}^{M} \|b_i\|^2 - 2 \sum_{i=1}^{M} (b_i,z) + M\|z\|^2 \\ &\geqslant \sum_{i=1}^{M} \|b_i\|^2 - 2M \left\| \sum_{i=1}^{M} \frac{b_i}{M} \right\| \cdot \|z\| + M\|z\|^2 \end{aligned} \tag{4.3}$$

The quadratic in $\|z\|$ is minimized at $\|z\| = \|y\|$, and the inequality is strict unless $z = y$. Moreover

$$\begin{aligned} G(z + \lambda(y-z),x) &= \sum_{i=1}^{M} \|b_i - z\|^2 - 2\lambda \sum_{i=1}^{M} (y - z, b_i - z) + \lambda^2 M\|y - z\|^2 \\ &= \sum_{i=1}^{M} \|b_i - z\|^2 - (2 - \lambda)\lambda M\|y - z\|^2 \\ &< \sum_{i=1}^{M} \|b_i - z\|^2 = G(z,x) \end{aligned} \tag{4.4}$$

since $z \neq y$ and $\lambda \in (0,2)$ imply that $\lambda(2-\lambda)\|y-z\|^2 > 0$.

Proposition 1: The sequence $\{x_n\}$ defined by the iteration

$$y_n = \frac{1}{M} \sum_{i=1}^{M} P_i x_n$$

$$x_{n+1} = x_n + \lambda_n (y_n - x_n), \qquad \lambda_n \in (0,2) \tag{4.5}$$

satisfies

$$F(x_{n+1}) < F(x_n) \quad \text{or} \quad x_{n+1} = x_n \ . \tag{4.6}$$

If the projection operators are single valued then $x_{n+1} = x_n$ implies that $x_{n+k} = x_n$ $\forall k$. If the sequence x_n has a limit point x at which the projections are continuous and

$$0 < \liminf \lambda_n \leqslant \limsup \lambda_n < 2$$

then x is a fixed point of the iteration.

Proof:

$$F(x_{n+1}) = \sum_{i=1}^{M} \|x_{n+1} - P_i x_{n+1}\|^2 \leqslant \sum_{i=1}^{M} \|x_{n+1} - P_i x_n\|^2$$

$$= G(x_n + \lambda_n (y_n - z_n), x_n) \tag{4.7}$$

$$\leqslant G(x_n, x_n) = F(x_n) \ .$$

From Lemma 1 the last inequality is strict if $x_n \neq y_n$. Therefore $F(x_{n+1}) = F(x_n)$ if and only if $y_n = x_n$ if and only if $x_{n+1} = x_n$ implies $y_{n+1} = y_n$, consequently $x_{n+k} = x_n$ for all k. Let $\tilde{x}$ be a limit point of x_n, if $\tilde{x}$ is not a fixed point then $\|\tilde{y} - \tilde{x}\|^2 = \varepsilon > 0$. By continuity of $\|\cdot\|$ and P_i at $\tilde{x}$, then $\lim x_{n_k} = \tilde{x}$ implies

a) $\lim F(x_{n_k}) = F(\lim x_{n_k}) = F(\tilde{x})$

b) $\lim y_{n_k} = \tilde{y}$.

Now choose x_N sufficiently close to $\tilde{x}$ such that

$$|F(x_N) - F(\tilde{x})| < \frac{1}{4}\alpha^2 M\varepsilon \tag{4.8}$$

and

$$\|y_N - x_N\|^2 \geq \frac{\varepsilon}{2} \tag{4.9}$$

where

$$\alpha = \min[\liminf \lambda_n,\ 2 - \limsup \lambda_n] \quad . \tag{4.10}$$

These inequalities imply

$$\begin{aligned} F(x_{N+1}) &\leq G(x_{N+1}, x_N) = G(x_N, x_N) - \lambda(2 - \lambda_N)M\|y_N - x_N\|^2 \\ &\leq F(x_N) - \frac{1}{2}\alpha^2 M\varepsilon \\ &< F(\tilde{x}) \quad . \end{aligned} \tag{4.11}$$

But as $F(x_n)$ is always decreasing, this implies the contradiction $\lim F(x_{n_k}) < F(\tilde{x})$.

Existence of limit points and convergence of the algorithm from an arbitrary x_0 cannot be guaranteed without further rather restrictive global conditions such as compactness of the sets B_i and continuity of the projections P_i. However examples showing failure of convergence after violation of these conditions are somewhat pathological; in practice, although the conditions may not be met, the algorithm almost always converges. Apparent failures are usually traced to ill conditioning not to violation or near violations of these conditions, therefore we assume henceforth the existence of limit points, leaving determination of sufficient conditions for this

existence to the analyst, and turn instead to an alternative characterization of these limit points and the algorithm itself.

4.2 *Successive Projections as Steepest Descent Algorithms*

To demonstrate that the generic algorithm of Proposition 1 is the standard steepest descent algorithm applied to $F(x)$ we need the following definition: Let $f:H\to\mathbb{R}$ be a real valued function on a Hilbert space H. Then the gradient $\nabla f(x)\in H$ of f at x is y if and only if y is the unique vector satisfying

$$\lim_{\|z\|\to 0}\left|\frac{f(x+z)-f(x)-(z,y)}{\|z\|}\right| = 0 . \tag{4.12}$$

The gradient $\nabla F(x)$ is the sum of the gradients of $\|x-P_i x\|^2$. We next require

Proposition 2. If there exists a $\lambda>1$ such that

$$P_A(P_A x+\lambda(x-P_A x)) = P_A x \tag{4.13}$$

and $P_A x$ is single valued, then

$$\nabla(\|x-P_A x\|^2) = 2(x-P_A x) . \tag{4.14}$$

Proof: It suffices to show that for some constant c the following inequality holds

$$\left|\|y-P_A y\|^2-\|x-P_A x\|^2-2(x-P_A x,y-x)\right| \leq c\|x-y\|^2 . \tag{4.15}$$

We first note the pair of inequalities.

$$\|y - P_A y\|^2 \leq \|y - P_A x\|^2 = \|y - P_A y\| + 2(y - P_A y, P_A y - P_A x) + \|P_A x - P_A y\|^2$$

$$\|y - P_A x\|^2 \geq \|y - P_A y\|^2 = \|y - P_A x\|^2 + 2(y - P_A x, P_A x - P_A y) + \|P_A x - P_A y\|^2 \quad (4.16)$$

and the resultant inequalities

$$2(y - P_A y, P_A y - P_A x) \geq - \|P_A y - P_A x\|^2 \quad (4.17)$$

$$2(y - P_A x, P_A x - P_A y) \leq - \|P_A y - P_A x\|^2 \quad (4.18)$$

We also require the equality

$$\|y - P_A y\|^2 - \|x - P_A x\|^2 - 2(y - x, x - P_A x)$$

$$= \|x - y\|^2 + \|P_A x - P_A y\|^2 + 2(y - x, P_A x - P_A y) + 2(x - P_A x, P_A x - P_A y) \quad (4.19)$$

We next bound $\|P_A x - P_A y\|$ in terms of $\|x - y\|$. By the hypothesis of the proposition

$$\|P_A x + \lambda(x - P_A x) - P_A y\|^2 \geq \|P_A x + \lambda(x - P_A x) - P_A x\|^2$$

$$= \lambda^2 \|x - P_A x\|^2 \quad (4.20)$$

which in turn implies

$$\|P_A x - P_A y\|^2 \geq 2\lambda(P_A x - x, P_A x - P_A y) \quad . \quad (4.21)$$

Now use of Equations (4.18) and (4.21) gives

$$\|P_A x - P_A y\|^2 \leq 2(P_A x - y, P_A x - P_A y)$$

$$= 2(P_A x - x, P_A x - P_A y) + 2(x - y, P_A x - P_A y)$$

$$\leq \frac{1}{\lambda} \|P_A x - P_A y\|^2 + 2\|x - y\| \|P_A x - P_A y\| \quad (4.22)$$

Consequently

$$\|P_A x - P_A y\| \leq \frac{2\lambda}{\lambda-1} \|x - y\| \tag{4.23}$$

which is the sought-for result. To establish the lower bound on the inequality in Eq. (4.15), reverse x and y in Eqs. (4.16) and (4.17), substitute the new Eq. (4.17) into Eq. (4.19) and use Eq. (4.23) to give

$$\|y - P_A y\|^2 - \|x - P_A x\|^2 - 2(y - x, x - P_A x) \geq -\frac{\lambda+1}{\lambda-1} \|x - y\|^2 . \tag{4.24}$$

For the upper bound, we note that the left hand side of Eq. (4.15) can be written as

$$-\|y - x\|^2 - \|P_A y - P_A x\|^2 - 2(x - y, P_A y - P_A x) - 2(y - P_A y, P_A y - P_A x)$$
$$- 2(x - y, y - x - (P_A y - P_A x)) \tag{4.25}$$

upon using Eq. (4.19) with x and y reversed. Further substitution of Eqs. (4.17) and (4.23) yields

$$\|y - P_A y\|^2 - \|x - P_A x\|^2 - 2(y - x, x - P_A x)$$
$$\leq -\|y - x\|^2 - \|P_A y - P_A x\|^2 + 2\|x - y\| \, \|P_A y - P_A x\| \tag{4.26}$$
$$+ \|P_A y - P_A x\|^2 + 2\|x - y\| [\|y - x\| + \|P_A y - P_A x\|] \leq \frac{9\lambda-1}{\lambda-1} \|x - y\|^2$$

which completes the proof.

This result is sufficient to show that

$$\nabla F(x) = 2 \sum_{i=1}^{M} (x - P_i x) = 2M(x - y) . \tag{4.27}$$

Hence the directions $y_n - x_n$ used in the algorithm of Proposition 1 are steepest descent directions for $F(x)$. Furthermore the result that, at a limit point $\tilde{x}, \tilde{y} = \tilde{x}$ shows that $\nabla F(\tilde{x}) = 0$ so $\tilde{x}$ is by definition a stationary point of $F(x)$.

4.3 *The Restricted Projection Algorithm*

In many reconstruction problems considerations such as computational complexity and firm restrictions on the functions g and G argue for the iterates being restricted to one particular set, B_M, say. We now show that the natural modification of the generic algorithm still produces a sequence with decreasing function values whose limit points are stationary points of $F(x)$, with x restricted to B_M.

Proposition 3. If the projections P_i are continuous and unique then the sequence $x_n \in B_M$ defined by

$$y_n = \frac{1}{(M-1)} \sum_{i=1}^{M-1} P_i x_n$$

$$x_{n+1} = P_M(x_n + \lambda_n(y_n - x_n)), \quad \lambda_n \in (0,1) \tag{4.28}$$

satisfies $F(x_{n+1}) < F(x_n)$ or $x_{n+k} = x_n$ for all k. Furthermore if $\tilde{x}$ is a limit point of x_n, then $\nabla F(\tilde{x})$ is normal to the set B_M.

Proof: The first step is to establish a bound on $(x_{n+1} - x_n,\ y_n - x_n)$. Let

$$z_n = x_n + \lambda_n(y_n - x_n) \tag{4.29}$$

then

$$\|x_{n+1} - z_n\|^2 = \|x_{n+1} - x_n\|^2 + \|x_n - z_n\|^2 + 2(x_{n+1} - x_n, x_n - z_n) \tag{4.30}$$

and

$$2(x_{n+1} - x_n, x_n - z_n) = 2\lambda_n(x_{n+1} - x_n, y_n - x_n)$$

$$= \|x_{n+1} - x_n\|^2 + \|x_n - z_n\|^2 - \|x_{n+1} - z_n\|^2 . \tag{4.31}$$

Since $P_M z_n = x_{n+1}$ we have that

$$2(x_{n+1} - x_n, y_n - x_n) \geq \frac{1}{\lambda_n} \|x_{n+1} - x_n\|^2 . \tag{4.32}$$

Therefore

$$\begin{aligned} F(x_{n+1}) &\leq \sum_{i=1}^{M-1} \|P_i x_n - x_{n+1}\|^2 \\ &= \sum_{i=1}^{M-1} \|P_i x_n - x_n\|^2 + 2 \sum_{i=1}^{M-1} (P_i x_n - x_n, x_n - x_{n+1}) + (M-1)\|x_{n+1} - x_n\|^2 \\ &= F(x_n) + (M-1)(x_n - x_{n+1}, 2y_n - x_n - x_{n+1}) \\ &= F(x_n) + (M-1)[2(x_n - x_{n+1}, y_n - x_n) + \|x_n - x_{n+1}\|^2] \\ &\leq F(x_n) + (M-1)\left(1 - \frac{1}{\lambda_n}\right)\|x_{n+1} - x_n\|^2 \end{aligned} \tag{4.33}$$

where the last line follows from Eq. (4.32); consequently

$$F(x_{n+1}) \leq F(x_n) \tag{4.34}$$

since $\lambda_n < 1$. Furthermore the inequality is strict unless $x_{n+1} = x_n$. Uniqueness of projections ensures that $x_{n+1} = x_n$ implies that $x_{n+k} = x_n \ \forall k$. Let $\tilde{x}$ be a limit point of x_n, then continuity of P_M implies that $\tilde{x} \in B_M$ and

$$\nabla F(\tilde{x}) = \sum_{i=1}^{M-1} 2(\tilde{x} - P_i \tilde{x}) = 2(M-1)(\tilde{x} - \tilde{y}) . \tag{4.35}$$

Since $\tilde{x}$ is a fixed point of the iteration (the same arguments as used in the proof of Proposition 1 shown this) then $P_M \tilde{y} = \tilde{x}$ and so the sphere $S \equiv \{z : \|z - \tilde{y}\| < \|\tilde{x} - \tilde{y}\|\}$ must satisfy $S \cap B_M = \phi$. But the sphere has a tangent plane at $\tilde{x}$, so B_M also has a tangent plane there; since $\tilde{y} - \tilde{x}$ is the normal to this tangent plane then $\tilde{y} - \tilde{x} \equiv \nabla F(\tilde{x})/2(M-1)$ is normal to B_M.

If the set B_M is convex, then the condition $\lambda_n \in (0,1)$ can be replaced by $\lambda_n \in (0,2)$. If B_M is a linear subspace then P_M is a linear map, so for any λ

$$P_M(x_n + \lambda(y_n - x_n)) = P_M x_n + \lambda P_M(y_n - x_n) \in B_M \quad . \tag{4.36}$$

Thus a unique search direction $P_M(y_n - x_n)$ is defined regardless of λ. A line search can be used to find the minimum of $F(P_M x_n + \lambda P_M(y_n - x_n))$ as a function of λ.

4.4 *Restricted Projections in Phase Retrieval*

The restricted projection algorithm is especially suitable for the phase retrieval problem for two reasons. The first reason stems from the requirement that $G(\omega) \in cI$; this set is a linear subspace and so is a natural choice for the set B_M of the previous subsection. The second reason is that such a choice is efficient. The problem as stated gives information on g and G only over the sets cI, we will show that the restricted projection algorithm with $B_M = cI$ requires knowledge of g_n and G_n at each iteration only over cI, even if the solution $g(v)$ is required over the entire real line. With the restricted projection algorithm, only the values $P_c g_n$ are used until the iterates converge. At this point $P_c \tilde{g}$ can be extrapolated to the whole line. The existence of efficient algorithms for phase retrieval has not always been realized, several authors, [14], have proposed schemes requiring storage of values of the iterates from *outside* the interval cI.

To show the efficiency of the algorithm, we present the details of the restricted projections arising in model problems I-III. Upon defining the set

$\mathcal{F}^{-1}A \equiv \{G: G = \mathcal{F}^{-1}g,\ g \in A\}$, it follows that $g \in A$ if and only if $\mathcal{F}^{-1}g \equiv G \in \mathcal{F}^{-1}A$. Since the Fourier transform preserves L^2 norms, $P_{\mathcal{F}^{-1}A}G$ is the closest point in $\mathcal{F}^{-1}A$ to G if and only if $\mathcal{F}P_{\mathcal{F}^{-1}A}G$ is the closest point in A to g, therefore

$$P_{\mathcal{F}^{-1}A}G = \mathcal{F}^{-1}P_A g \quad . \tag{4.37}$$

$P_A g$ is easily shown to be

$$\begin{aligned} (P_A g)(v) &= m(v)e^{i \arg g(v)}, \quad v \in cI \\ &= g(v) \qquad\qquad , \quad \text{otherwise} \end{aligned} \tag{4.38}$$

Therefore, after noting that $P_c P_A = P_c P_A P_c$ and restricting G to cI,

$$P_A g = g - P_c g + P_c P_A P_c g \tag{4.39}$$

and

$$\begin{aligned} P_c P_{\mathcal{F}^{-1}A}G &= P_c \mathcal{F}^{-1}(\mathcal{F}G - P_c \mathcal{F}G + P_c P_A P_c \mathcal{F}G) \\ &= G - P_c \mathcal{F}^{-1}P_c \mathcal{F}P_c G + P_c \mathcal{F}^{-1}P_c P_A P_c \mathcal{F}P_c G \quad . \end{aligned} \tag{4.40}$$

As the remaining sets B_i are by definition contained in cI, it follows that $P_c P_{B_i} = P_{B_i}$. For the record these projections are

$$P_{B_1}G = P_c G \tag{4.41}$$

$$\begin{aligned} (P_{B_2}G)(\omega) &= (\text{Re } G)(\omega), \quad \text{if} \quad (\text{Re } G)(\omega) > 0, \ \omega \in cI \\ &= 0 \qquad\qquad , \quad \text{otherwise} \end{aligned} \tag{4.42}$$

$$\begin{aligned} (P_{B_3}G)(\omega) &= n(\omega)e^{i \arg G(\omega)}, \quad \text{if} \quad \omega \in cI \\ &= 0 \qquad\qquad\qquad , \quad \text{otherwise} \quad . \end{aligned} \tag{4.43}$$

For typographic convenience we will let $P_{\mathcal{F}^{-1}A} \to P_1$, $P_{B_2} \to P_2$ and $P_{B_3} \to P_3$; also we will denote the restricted projection algorithm using these projections and the set $B_M = cI$ by RP. Therefore the iterates x_n become iterates $G_n \in cI$. Further we shall denote the search directions $y_n - x_n$ by H_n. For RP, H_n is given by

$$H_n = \frac{1}{M} \sum_{i=1}^{M} (P_c P_i G_n - G_n) \quad . \tag{4.44}$$

The calculation of H_n requires values of g_n and G_n only over cI after use of Eq. (4.40) for $P_c P_1$.

4.5 *Ill Conditioning and Line Searches*

The geometric view of the RP algorithm developed in this section sheds new light on some of the problems encountered in previous use of projection algorithms, [7], in particular that of constant but small decrements in $F(G_n)$. This may be due to local ill conditioning of the problem; with the geometric interpretation that the sets $\mathcal{F}^{-1}A$ and B_i are nearly parallel at G_n, either intersecting at a very acute angle or failing to intersect at all. Figure 1 shows simple two-dimensional examples of such problems. In both cases shown in Fig. 1, H_n is very small so that if $\lambda_n \in (0,2)$ then G_n and G_{n+1} are nearly coincident even though G_n is far from the true minimum.

The relation established between projection methods and steepest descent methods opens up a wide range of algorithms already developed for gradient methods. In particular, we note that calculation of $F(G_n)$ requires evaluation of $P_c P_i G_n$ so we may simultaneously calculate

$$\frac{d}{d\lambda} F(G_n + \lambda H_n) = (\nabla F(G_n + \lambda H_n), H_n)$$

$$= 2 \sum_{i=1}^{M} (G_n + \lambda H_n - P_i(G_n + \lambda H_n), H_n)$$

$$= 2 \sum_{i=1}^{M} (G_n + \lambda H_n - P_c P_i(G_n + \lambda H_n), H_n) \qquad (4.45)$$

because

$$H_n = P_c H_n, \qquad (G, P_c H) = (P_c G, H) \quad . \qquad (4.46)$$

Therefore a quadratic [15] or cubic [16] line search routine will improve convergence beyond simple variation of λ_n in $(0,2)$.

5. SECOND ORDER ITERATIVE METHODS

The iterative algorithms for minimization of $F(G)$ discussed in the previous section were developed from first order truncations, either approximating $F(G)$ by its gradient $\nabla F(G)$ or taking pointwise approximations to the sets A and B_i. In a similar vein we now construct second order algorithms by approximations involving the Hessian $\mathcal{H}(G)$ of $F(G)$ and affine approximations to the sets. A succession of such algorithms is constructed, increasing in accuracy and in complexity up to the standard Newton's method for minimization of $F(G)$. In each algorithm a linear subproblem must be solved at each stage; typically the subproblem is ill-conditioned and requires a filtered inversion. The geometric viewpoint shows that the system can be taken to be a linear least squares problem whose solution is the closest point to a collection of affine subspaces approximating the sets A and B_i. We then demonstrate that in this new form the system may be preconditioned so that a filtered inversion is possible without having to calculate an expensive singular value decomposition at each stage.

Unlike pointwise approximations, the affine approximations to the sets are not necessarily contained within the sets, therefore we cannot guarantee that $F(G_n + \lambda H_n) < F(G_n)$ for $\lambda \in (0,2)$. However, it is often possible to show that the search directions H_n generated at each iteration are descent directions, i.e. $(d/d\lambda)F(G_n + \lambda H_n)_{\lambda=0} < 0$. In the interests of brevity only the form of the algorithms is presented here, proof of these and other similar results are left to the reader.

5.1 *The Partial Affine Algorithm*

As an introduction to the use of affine approximations we present a simple extension of RP. In RP at the n-th iteration the set A is approximated by the point $P_A g_n$ so that the restricted projection $P_c P_1 G_n$ is given by Eq. (4.40). However membership of g in A is determined only by the modulus values of g across cI; outside of this interval there are no restrictions. Therefore the affine subspace

$$A_n \equiv \{h : h = P_c P_A g_n + (g - P_c g),\ g \in L^2(\mathbb{R})\} \tag{5.1}$$

contains the point $P_A g_n$ but is contained in A. So for any function $G \in cI$

$$\|G - P_1 G\| \leq \|G - P_{\mathcal{F}^{-1} A_n} G\| = \|G - \mathcal{F}^{-1} P_{A_n} g\| \quad . \tag{5.2}$$

Use of A_n and the points $P_i G_n$ as approximations to sets gives an approximation F_n to F such that

$$\begin{aligned} F_n(G) &= \|P_c P_A g_n - P_c \mathcal{F} P_c G\|^2 + \sum_{i=2}^{M} \|P_i G_n - G\|^2 \\ F(G) &\leq F_n(G), \qquad F(G_n) = F_n(G_n) \end{aligned} \tag{5.3}$$

$F_n(G)$ is minimized at the point L_n satisfying the normal equations

$$(P_c \mathcal{F}^{-1} P_c \mathcal{F} P_c + (M-1)I) L_n = P_c \mathcal{F}^{-1} P_c P_A g_n + \sum_{i=2}^{M} P_i G_n \tag{5.4}$$

giving a search direction $H_n = L_n - G_n$. If $M > 1$, Eq. (5.4) can be solved by any regular linear equations package since the eigenvalues of

$(P_c \mathcal{F}^{-1} P_c \mathcal{F} P_c + (M-1)I)$ are all greater than unity and less than M. If $M = 1$, then minimization of $F_n(G)$ reduces to the least squares problem

$$\min_{G \in cI} \| P_c \mathcal{F} P_c G - P_c P_A g_n \|^2 \tag{5.5}$$

which is the focus of the previous paper. In this case the ill conditioning of the finite Fourier transform $P_c \mathcal{F} P_c$ forces the use of a filtered approximation $\tilde{L}_n$ to L_n in which $\tilde{L}_n$ is the projection of L_n onto the span of the first few eigenfunctions of $P_c \mathcal{F} P_c$. Since the same linear operator appears at each iteration, the eigendecomposition need be calculated only once. As the range of the projection remains unchanged, an extra global restriction on the iterates to this span is implicitly imposed.

Henceforth the algorithm with search direction given by Eq. (5.4) will be termed the PA algorithm.

5.2 *The Gauss-Newton Algorithm*

Having introduced the geometric viewpoint of algorithms as based upon set approximations, we now look at some of the standard nonlinear least squares algorithms in this context. For solution of the model problem

$$\min_{\hat{x} \in \mathbb{R}^N} F(\hat{x}) = \min_{\hat{x} \in \mathbb{R}^N} \sum_{i=1}^{M} f_i^2(\hat{x}) \tag{5.6}$$

a popular choice of iterative algorithm is the Gauss-Newton method [17] in which at the n-th iteration the search direction $\hat{z}_n$ is given by the least squares minimum norm solution to the linear system

$$\hat{J}(\hat{x}_n)\hat{z}_n = -\hat{f}(\hat{x}_n) \tag{5.7}$$

where $\hat{f}(\hat{x}_n)$ is the vector in $\mathbb{R}^M$ with components $f_i(\hat{x}_n)$ and $\hat{J}(\hat{x}_n)$ is the $M \times N$ matrix with entries

$$\hat{J}_{ij}(\hat{x}_n) \equiv \frac{\partial}{\partial x_j} f_i(\hat{x}_n) \quad . \tag{5.8}$$

In the present paper $x \equiv G \in L^2(\mathbb{R})$ and $f_i(x) = \|x - P_i x\|$; the gradient of $f_i(x)$ as defined by Eq. (4.12) is

$$(\nabla f_i)(x) = (x - P_i x)/\|x - P_i x\| \tag{5.9}$$

so that the Jacobian $\mathcal{J}$ is now an operator from $L^2(\mathbb{R})$ into $\mathbb{R}^M$ defined by

$$[\mathcal{J}(x)z]_i = (x - P_i x, z)/\|x - P_i x\| \quad . \tag{5.10}$$

Therefore z_n is now the least squares, minimum norm solution to the underdetermined system, Eq. (5.7), with $\mathcal{J}(x_n)$ replacing $\hat{J}(\hat{x}_n)$; such a z_n gives a descent direction for $F(x)$.

We next elucidate the affine approximation implicit in this algorithm. Since $P_i x_n$ is the unique closest point to x_n from B_i, it follows that if $S_i(x_n)$ is the sphere

$$S_i(x_n) \equiv \{z : \|x_n - z\| \leq \|x_n - P_i x_n\|\} \tag{5.11}$$

then $B_i \cap S_i(x_n) = P_i x_n$. At $P_i x_n$ the sphere has a well defined tangent plane

$$T_i(x_n) \equiv \{P_i x_n + z : (x_n - P_i x_n, z) = ((\nabla f_i)(x_n), z) = 0\} \quad . \tag{5.12}$$

This must also be locally a supporting hyperplane for B_i at $P_i x_n$; Fig. 2 shows a simple two dimensional example. $T_i(x_n)$ can be taken as an affine approximation to B_i at the n-th iteration, in which case the least squares, minimum norm solution to Eq. (5.4) is the closest point in $\cap_{i=1}^{M} T_i(x_n)$ to x_n. Since $T_i(x_n)$ has codimension one, this intersection is always nonempty, being in fact an infinite dimensional linear subspace.

Although we bring this algorithm to the readers attention we did not implement it for two reasons. The first is that the subspaces $T_i(x_n)$ are poor approximations to the sets B_i, containing no further information than that already available in the projections $P_i x_n$. Second for our model problems $M \leq 3$, so that Eq. (5.7) is of very small dimensions. Therefore the search directions z_n are not likely to be very different from those of RP or PA.

5.3 *Newton's Algorithm*

The standard second order Newton's algorithm for Eq. (5.6) is based on the approximation of $F(z)$ near $x = x_n$ by a truncated Taylor series expansion

$$F(x) \approx F(x_n) + \nabla F(x_n)(x-x_n) + \frac{1}{2}(x-x_n)\nabla^2 F(x_n)(x-x_n) \quad . \tag{5.13}$$

If $\nabla^2 F(x_n)$ is a positive definite matrix, then the approximation is minimized at the point $y_n = z_n + x_n$, where z_n is the solution of

$$\nabla^2 F(x_n) z = -\nabla F(x_n) \quad . \tag{5.14}$$

In phase retrieval problems this linear system is ill conditioned and the Hessian often has negative eigenvalues. Therefore to ensure that a descent direction is chosen for which Eq. (5.13) is a good approximation, an eigendecomposition of the Hessian must be calculated and a filtered solution z_n used. This approach must be taken if Eq. (5.14) is solved directly. However, as we will show, a geometric interpretation allows the equation to be rewritten in a form in which filtering can be performed without the cost of an eigendecomposition at each iteration.

We begin by noting that the quantities $\nabla F(x_n)$ and $\nabla^2 F(x_n)$ are given by

$$\nabla F(x_n) = 2 \sum_{i=1}^{M} (x_n - P_i x_n) \tag{5.15}$$

$$\nabla^2 F(x_n) = 2 \sum_{i=1}^{M} (I - \mathscr{H}_i(x_n)) \tag{5.16}$$

where the operators $\mathscr{H}_i(x_n)$ are the Hessians of the functions $f_i(x) = \|x - P_i x\|^2$ evaluated at x_n. They are defined by

$$\begin{aligned} &\mathscr{H}_i(x_n) : L^2(\mathbb{R}) \to L^2(\mathbb{R}) \\ &\mathscr{H}_i(x_n) y = \lim_{\alpha \to 0} \frac{P_i(x_n + \alpha y) - P_i(x_n)}{\alpha} \end{aligned} \tag{5.17}$$

Conditions on the set B_i that guarantee existence and boundedness of the operator $\mathscr{H}_i$ are quite complicated [18]. For the purposes of this paper we shall assume that $\mathscr{H}_i(x_n)$ exists. The only further properties we require are that $\mathscr{H}_i$ is symmetric and that

$$\mathcal{H}_i(x_n)(x_n - P_i x_n) = 0 \quad . \tag{5.18}$$

The Hessian implicitly defines a first order approximation to the projection operator P_i at x_n by

$$P_i x \approx P_i x_n + \mathcal{H}_i(x_n)(x - x_n) \tag{5.19}$$

which in turn defines an associated affine approximation $U_i(x_n)$ to the set B_i at $P_i(x_n)$

$$U_i(x_n) \equiv \{P_i x_n + z : z \in \text{Range}\, \mathcal{H}_i(x_n)\} \quad . \tag{5.20}$$

$U_i(x_n)$ has more structure than the previous approximation $T_i(x_n)$ because $\mathcal{H}_i(x_n)$ is not just a projection operator with range $U_i(x_n) - P_i x_n$ but is also a contraction (or expansion) mapping about $P_i x_n$ *depending on* whether B_i is locally convex (or concave) in particular directions at $P_i x_n$. Figure 3 shows a two-dimensional example of $U_i(x_n)$ for B_i a sphere; $\mathcal{H}_i(x)$ is obviously a contraction mapping. Therefore at the n-th iteration we choose a search direction $z_n = y_n - x_n$ where y_n minimizes the quadratic approximation

$$F(y) \approx \sum_{i=1}^{M} \| y - P_i x_n - \mathcal{H}_i(x_n)(y - x_n) \|^2 \tag{5.21}$$

to $F(x)$. The geometric view of y_n is that it is the "closest point" to the collection of subspaces $U_i(x_n)$ with the distance measured, not in the standard L^2 metric, but as a sum of new metrics. Each new metric reflects the degree of curvature of the set B_i at $P_i x_n$, and therefore the accuracy of the affine approximation to B_i at $P_i x_n$.

The direction z_n of Eq. (5.21) is equivalently described as the least squares solution to the block system

$$\begin{vmatrix} I - \mathscr{H}_1(x_n) \\ \vdots \\ I - \mathscr{H}_M(x_n) \end{vmatrix} z = \begin{vmatrix} -(x_n - P_1 x_n) \\ \vdots \\ -(x_n - P_M x_n) \end{vmatrix} . \tag{5.22}$$

Equations (5.22) and (5.14) are not the same, so that the search directions z_n are different. However, the equations are sufficiently close in form to show how Eq. (5.14) can be recast in a block form more suitable for numerical computation. To accomplish this, we note that the normal equations for Eq. (5.22) are

$$\left(\sum_{i=1}^{M} (I - \mathscr{H}_i(x_n))^2\right) z = -\sum_{i=1}^{M} (I - \mathscr{H}_i(x_n))(x_n - P_i x_n)$$

$$= -\sum_{i=1}^{M} (x_n - P_i x_n) \tag{5.23}$$

after use of Eq. (5.18). If the operators $I - \mathscr{H}_i(x_n)$ are positive definite, then they possess a positive definite square root $(I - \mathscr{H}_i(x_n))^{1/2}$ which by Eq. (5.18) satisfies

$$(I - \mathscr{H}_i(x_n))^{1/2}(x_n - P_i x_n) = x_n - P_i x_n \quad . \tag{5.24}$$

Therefore Eq. (5.14) can be rewritten as

$$\left[\sum_{i-1}^{M} ((I - \mathscr{H}_i(x_n))^{1/2})^2\right] z = -\sum_{i=1}^{M} (I - \mathscr{H}_i(x_n))^{1/2}(x_n - P_i x_n) \tag{5.25}$$

which in turn is recognizable as the normal equations associated with the block system

$$\begin{vmatrix} (I - \mathscr{H}_1(x_n))^{1/2} \\ \vdots \\ (I - \mathscr{H}_M(x_n))^{1/2} \end{vmatrix} \hat{z} = \begin{vmatrix} -(x_n - P_1 x_n) \\ \vdots \\ -(x_n - P_M x_n) \end{vmatrix} . \tag{5.26}$$

Equation (5.26) has a geometric interpretation, its solution is the solution to an approximate minimization of $F(x)$ similar to that of Eq. (5.21) using the same affine approximation $U_i(x_n)$ to B_i but different metrics.

5.4 *Ill Conditioning and Filtering*

We noted at the end of Section 4 that line searches could overcome some simple cases of ill conditioning. We now consider more complicated cases and ways to alleviate them.

The first comes from the fact that $L^2(cI)$ is not one-dimensional, as portrayed in Fig. 1, so that the valleys of the surface $T \equiv \{(F(G),G):G \in cI\}$ will not typically be straight but rather will be curving through space in much the same fashion as the classical test case for optimization algorithms, the Rosenbrock function [16]. Experience has shown that second order methods significantly outperform gradient methods in such examples, but that such features can easily be sufficiently ill conditioned to defeat even second order algorithms.

The second arises from the existence of the subspace S described in Section 3 which is, for practical purposes, the null space of $P_c \mathscr{H} P_c$. If a

significant component of the search direction lies within this subspace then a relatively large step can be taken for a very small decrement in $F(G)$. This may lead to "zig-zags" in which steps are taken backwards and forwards through S so that the iterates do not appear to converge and yet $F(G_n)$ is decreasing, albeit very slowly. The presence of S also implies that the Hessian $\nabla^2 F(G_n)$ will have very small eigenvalues corresponding to directions in S (particularly in Hessians from model problem I) so that Eq. (5.14) is difficult to solve numerically.

These problems can be overcome by calculation of the eigendecomposition of $\nabla^2 F(G_n)$, filtering the eigenvalues (which are real since the Hessian is symmetric) by choice of a cutoff parameter ε_3 and use of the filter

$$\begin{aligned} f(\sigma) &= \sigma \quad , \quad \text{if} \quad |\sigma| \geq \varepsilon_3 \\ &= 0 \quad , \quad \text{otherwise} \ , \end{aligned} \tag{5.27}$$

then finding the minimum norm least squares solution z_n of the resulting filtered version of Eq. (5.14). This scheme produces a search direction in which $F(G)$ should vary moderately rapidly because directions corresponding to small eigenvalues, and thus slowly varying $F(G)$, have been filtered out. We denote such a filtered Newtonian algorithm by FN. The resulting direction z_n is not necessarily a descent direction ($\nabla^2 F(G_n)$ may have large negative eigenvalues) so we shall also consider a variant of FN using the filter

$$\begin{aligned} f(\sigma) &= \sigma \quad , \quad \text{if} \quad \sigma \geq \varepsilon_3 \\ &= 0 \quad , \quad \text{otherwise} \ . \end{aligned} \tag{5.28}$$

The resulting algorithm is termed FNP.

The problem of global ill conditioning, in particular the presence of multiple minima due to possible nonuniqueness of the phase retrieval problem, *cannot* be addressed by similar modifications to these undertaken for local ill conditioning simply because the algorithms are derived from local analysis.

5.5 *Block Filtering*

Having shown the need for filtering, we now show that Eq. (5.14) can be solved by a filtered solution of Eq. (5.26) in which each block is prefiltered and a standard least squares algorithm then applied (avoiding the cost of a singular value decomposition of the block matrix). We begin by noting that if our algorithm is to be efficient then we must solve

$$P_c \nabla^2 F(G_n) H = -P_c \nabla F(G_n) \tag{5.29}$$

rather than Eq. (5.14), so that values only over the interval cI are used. (For this reason, Eq. (5.29) rather than Eq. (5.14) is also used in FN and FNP). We also note that the Hessians are not complex valued linear operators since it will be obvious that for an arbitrary complex number α

$$\mathscr{H}_i(G_n)(\alpha H) \neq \alpha(\mathscr{H}_i(G_n)H) \quad . \tag{5.30}$$

Rather they are linear operators on the pair of real functions $(\text{Re } H)(\omega)$ and $(\text{Im } H)(\omega)$.

The ability to do block prefiltering will depend upon the special form of the Hessian in Eq. (5.26). For the operators $\mathscr{H}_2$ and $\mathscr{H}_3$ associated with P_2 and P_3 it is obvious that $P_c \mathscr{H}_i \equiv \mathscr{H}_i$. In addition, simple calculations establish

$$\mathcal{H}_2(G_n)\begin{vmatrix}(\mathrm{Re}\ H)(\omega)\\(\mathrm{Im}\ H)(\omega)\end{vmatrix} = \begin{vmatrix}(\mathrm{Re}\ H)(\omega)\\0\end{vmatrix}, \quad \text{if} \quad (\mathrm{Re}\ G_n)(\omega) > 0, \ \omega \in cI$$

$$= \begin{vmatrix}0\\0\end{vmatrix}, \quad \text{otherwise} \tag{5.31}$$

and

$$\mathcal{H}_3(G_n)\begin{vmatrix}(\mathrm{Re}\ H)(\omega)\\(\mathrm{Im}\ H)(\omega)\end{vmatrix} = \frac{n(\omega)}{|G_n(\omega)|}\begin{vmatrix}\sin^2\theta & -\sin\theta\cos\theta\\-\sin\theta\cos\theta & \cos^2\theta\end{vmatrix}\begin{vmatrix}(\mathrm{Re}\ H)(\omega)\\(\mathrm{Im}\ H)(\omega)\end{vmatrix} \tag{5.32}$$

where

$$\theta = \arg G_n(\omega), \qquad \omega \in cI\ . \tag{5.33}$$

The operators $I - \mathcal{H}_i$ can be represented by 2×2 block diagonal operators; at each point ω the blocks are

$$[I - \mathcal{H}_2(G_n)](\omega) = \begin{vmatrix}0 & 0\\0 & 1\end{vmatrix}, \quad \text{if} \quad (\mathrm{Re}\ G_n)(\omega) > 0$$

$$= \begin{vmatrix}1 & 0\\0 & 1\end{vmatrix}, \quad \text{otherwise} \tag{5.34}$$

$$[I - \mathcal{H}_3(G_n)](\omega) = \begin{vmatrix}\sin\theta & \cos\theta\\-\cos\theta & \sin\theta\end{vmatrix}\begin{vmatrix}1 - \frac{n(\omega)}{|G_n(\omega)|} & 0\\0 & 1\end{vmatrix}\begin{vmatrix}\sin\theta & -\cos\theta\\\cos\theta & \sin\theta\end{vmatrix}. \tag{5.35}$$

The spectrum of each operator $I - \mathcal{H}_i$ is now seen to be simply the union over ω of the spectrum of each block. For $I - \mathcal{H}_2$ this gives a spectrum

consisting only of the points $\{0,1\}$, so $I - \mathcal{H}_2$ is its own square root. Therefore the only prefiltering that needs to be done to $(I - \mathcal{H}_2)^{1/2}$ is to eliminate all equations for $(\text{Re } H)(\omega)$ that correspond to zeros in the block spectra. For $I - \mathcal{H}_3$, the spectrum will take on a range of values. Some of these values may be small or negative if for some ω the quantity $(1 - n(\omega)/|G_n(\omega)|)$ is small or negative. Consequently $(I - \mathcal{H}_3)^{1/2}$ either is not well defined (i.e. has imaginary eigenvalues) or is ill conditioned, or is both. To prefilter the block $(I - \mathcal{H}_3)^{1/2}$ we therefore choose a parameter $\varepsilon_3 > 0$ and *one* of the following filters

$$\begin{aligned} f(\sigma) &= |\sigma|^{1/2} \operatorname{sgn} \sigma, && \text{if } |\sigma| \geq \varepsilon_3 \\ &= 0, && \text{otherwise} \end{aligned} \tag{5.36}$$

$$\begin{aligned} f(\sigma) &= \sigma^{1/2}, && \text{if } \sigma \geq \varepsilon_3 \\ &= 0, && \text{otherwise} . \end{aligned} \tag{5.37}$$

The prefiltered block is then defined as

$$(I - \mathcal{H}_3(G_n))_f^{1/2}(\omega) = \begin{vmatrix} \sin\theta & \cos\theta \\ -\cos\theta & \sin\theta \end{vmatrix} \begin{vmatrix} f\left(1 - \dfrac{n(\omega)}{|G_n(\omega)|}\right) & 0 \\ 0 & 1 \end{vmatrix} \begin{vmatrix} \sin\theta & -\cos\theta \\ \cos\theta & \sin\theta \end{vmatrix} . \tag{5.38}$$

It is reduced in size by eliminating those equations for $\sin\theta(\text{Re } H)(\omega) - \cos\theta(\text{Im } H)(\omega)$ that are associated with elements in the spectrum that are filtered out.

From Eq. (4.39) $\mathcal{H}_1$ has the form

$$\mathcal{H}_1(G_n) = I - \mathcal{F}^{-1} P_c \mathcal{F} P_c + \mathcal{F}^{-1} P_c \mathcal{H}_A(g_n) P_c \mathcal{F} P_c \tag{5.39}$$

so that

$$(I - P_c \mathcal{H}_1)(G_n) = P_c \mathcal{F}^{-1} P_c (I - \mathcal{H}_A(g_n)) P_c \mathcal{F} P_c \tag{5.40}$$

where $(I - \mathcal{H}_A(g_n))$ is in block diagonal form over cI:

$$[I - \mathcal{H}_A(g_n)](v) = \begin{vmatrix} \sin\phi & \cos\phi \\ -\cos\phi & \sin\phi \end{vmatrix} \begin{vmatrix} 1 - \frac{m(v)}{|g_n(v)|} & 0 \\ 0 & 1 \end{vmatrix} \begin{vmatrix} \sin\phi & -\cos\phi \\ \cos\phi & \sin\phi \end{vmatrix} \tag{5.41}$$

with

$$\phi \equiv \arg g_n(v), \qquad v \in cI \tag{5.42}$$

Unfortunately $I - P_c \mathcal{H}_1$ does not possess an obvious symmetric square root. However an asymmetric root can be found by noting that

$$(I - P_c \mathcal{H}_1(G_n))H = G_n - P_c P_1 G_n \tag{5.43}$$

is the normal equation associated with the least square problem

$$\min_H \| (I - \mathcal{H}_A(g_n))^{1/2} P_c \mathcal{F} P_c H - P_c P_A P_c \mathcal{F} P_c G_n \|^2 \tag{5.44}$$

so that we may take for the block $(I - P_c \mathcal{H}_1)^{1/2}$ the composite operator $(I - \mathcal{H}_A)^{1/2}(P_c \mathcal{F} P_c)$ and prefilter each component. The filtered component $(I - \mathcal{H}_A)_f^{1/2}$ is calculated in the same fashion as $(I - \mathcal{H}_3)_f^{1/2}$; the component $P_c \mathcal{F} P_c$ requires further manipulations before prefiltering. Let $P_c \mathcal{F} P_c$ have an eigendecomposition $U \Sigma U^+$, then as U is a unitary operator the solution H_n to a prefiltered Eq. (5.26) is given by $H_n = U L_n$, where L_n is the least

squares solution to

$$\begin{vmatrix} (I - \mathcal{H}_A(g_n))_f^{1/2} U\Sigma \\ (I - \mathcal{H}_i(G_n))_f^{1/2} U \end{vmatrix} L = \begin{vmatrix} P_c P_A P_c \mathcal{F} P_c G_n \\ P_i G_n - G_n \end{vmatrix} \quad (5.45)$$

where Σ may be filtered using Eqs. (5.36) or (5.37).

The reason for choosing to solve for L_n instead of H_n lies in the choice of the pointwise discretization used in numerical calculations and in the form of the IMSL subroutine LLSQF used to solve the discretized Eq. (5.45). This subroutine uses an adaptive QR algorithm, at each stage a column is chosen from the block matrix and incorporated in the QR factorization calculated at the previous stage. A running estimate is kept on the condition number of R, if this estimate exceeds a use defined parameter, ATOL, then the routine halts and computes a minimum norm least squares solution using the most recent QR factorization and also sets components in L corresponding to unused columns to zero. For a solution H this corresponds to setting point values of $H(\omega)$ to zero; for L it is setting components of L in an eigenfunction expansion to zero. Our analysis of ill conditioning indicates that such components are likely to be zero anyway, so we believe Eq. (5.45) to be the preferred form for solution.

We shall term the algorithm based on Eq. (5.45) with the filter given by Eq. (5.24) the SQFN algorithm, and that based on Eq. (5.45) with the filter given by Eq. (5.37) the SQFNP algorithm. SQFN incorporates more information about the Hessian at each step, but SQFNP gives a guaranteed descent direction H_n at each step.

6. NUMERICAL RESULTS

We now present results for the solution of the discretized model problems using the various algorithms so far proposed and show the discrete approximate solutions $\tilde{\hat{G}}$ generated. The main measure of performance used in this section will be the number of iterations $\tilde{n}$ required by the algorithm to reduce the function $F(\hat{G}_{\tilde{n}})$ to below a prescribed value. Since we are interested more in the comparative than absolute behaviour of the algorithms we take as the benchmark case the RP algorithm and consider the quantity $\tilde{n}/\check{n}$ where $\check{n}$ is the number of iterations required by RP to reduce $F(\hat{G}_{\check{n}})$ to below a prescribed value.

The ill conditioned nature of the problem means that this is not the best of measures. Difficulties are encountered through behaviour such as the erratic decrease of $F(\hat{G}_n)$; often a succession of iterates produce little change in $F(\hat{G}_n)$ and then a sudden decrease is recorded. In some algorithms the iterates may converge to a particular value $\tilde{\hat{G}}$ whereas in others they fail to converge so that $F(\hat{G}_n)$ and $F(\tilde{\hat{G}})$ can be considerably different. For these reasons the results reported here are intended as a guide to the behaviour of algorithms on real problems rather than a prediction. However, poor though the measure is, there is none other available for use that does not require prior knowledge of the solution.

6.1 *Discretization, Test Functions and Algorithms*

The choice of discretization in these calculations was determined by the condition that it give an accurate, efficient approximation to the

finite Fourier transform and that the discretized projections be easily calculated. Therefore a discretization based on N point Gaussian quadrature was chosen; the previous paper shows its accuracy and efficiency and its pointwise nature allows easy evaluation of projections. The finite dimensional problem now involves vectors $\hat{g}, \hat{G} \in C^N$ whose elements g_k, G_k are to be approximations to the function values $g(\rho_k), G(\rho_k)$ at the abscissae ρ_k of the N point Gaussian quadrature rule on cI. Vectors $\hat{m}, \hat{n} \in \mathbb{R}^N$ are formed with components $m_k = m(\rho_k)$, $n_k = n(\rho_k)$ where $m(v)$ and $n(w)$ are the known moduli. Matrices $\hat{W}, \hat{F} \in C^{N \times N}$ are constructed with $\hat{W}$ being a diagonal matrix whose k-th entry is the weight ω_k of the quadrature rule and with $\hat{F}$ having entries $F_{k\ell} = \ell^{2\pi i \rho_k \rho_\ell}$. This construction gives as the discretized model problems

Find vectors $\hat{g}$ *and* $\hat{G}$ *such that* $\hat{g} = \hat{F}\hat{W}\hat{G}$ *and*

I: $|g_k| = m_k$

II: $|g_k| = m_k, \quad G_\ell \geq 0$

III: $|g_k| = m_k, \quad |G_\ell| = n_\ell$

where $\hat{m}$ and $\hat{n}$ are specified.

Since the metric for all analysis so far has been $\| \ \|_2$ the numerical calculations were done on a rescaled version of the equation $\hat{g} = \hat{F}\hat{G}$,

$$(\hat{W}^{1/2}\hat{g}) = (\hat{W}^{1/2}\hat{F}\hat{W}^{1/2})(\hat{W}^{1/2}\hat{G}) \tag{6.1}$$

With this rescaling the Euclidean norms of the vectors $\hat{W}^{1/2}\hat{G}$ and $\hat{W}^{1/2}\hat{g}$ are now good approximations to the norms $\|G\|_2$ and $\|g\|_2$, and $\hat{W}^{1/2}\hat{F}\hat{W}^{1/2}$ is a good approximation to $P_c \mathcal{F} P_c$ in the operator norm induced by $\| \cdot \|_2$.

The two test functions chosen for the model problems were

$$G^1(\omega) = e^{\omega}, \qquad G^2(\omega) = e^{\frac{3}{4}\pi i\left[P_3\left(\frac{\omega}{c}\right) + P_4\left(\frac{\omega}{c}\right)\right]} \tag{6.2}$$

where $P_i(\omega)$ is the i-th monic Legendre polynomial. The Fourier transform $g^1(v)$ of $G^1(\omega)$ is

$$g^1(v) = 2\,\frac{\sin(2\pi v - i)c}{(2\pi v - i)}\ , \tag{6.3}$$

$g^2(v)$ has no closed form and was calculated numerically. Both functions were used as test cases for model problem I, G^1 and g^1 as a test case for problem II and G^2 and g^2 as a test case for problem III. The discrete approximations to $G^i(\omega)$ and $g^i(v)$ are denoted by $\hat{G}^i$ and $\hat{g}^i$.

All numerical tests used the following basic algorithm

1: Choose an initial guess G_0

2: Given iterates $\hat{g}_n$ and $\hat{G}_n$ compute a search direction $\hat{H}_n$

3: Calculate a step length λ_n and new iterates

$$\hat{G}_{n+1} = \hat{G}_n + \lambda_n \hat{H}_n\ , \qquad \hat{g}_{n+1} = \hat{F}\hat{W}\hat{G}_{n+1}$$

4: Iterate steps 2 and 3 until the convergence criteria are satisfied.

The convergence criteria used in all calculations were

$$F(\hat{G}_n) < 10^{-5} \quad \text{or} \quad \sum_{k=0}^{2} \|\lambda_{n-k}\hat{H}_{n-k}\| < 2 \times 10^{-3} \tag{6.4}$$

together with an upper limit N_{max} on the number of iterations. The sum of the last three steplengths, rather than $\|\lambda_n \hat{H}_n\|$ alone, was chosen as in ill conditioned minimization problems "stop-start" behaviour is often noticed.

That is a large step often followed by one or two small steps after which another large step is taken. This feature was often observed in the use of any second order algorithm, differences in magnitude of successive step lengths by factors greater than 100 occurred fairly frequently.

Three different forms of initial guess were used

1. $\hat{G}_0 = 0$
2. $G_{0\ell} = (1-\varepsilon_1)(1+\varepsilon_1 r_\ell) G_\ell^i$
3. $G_{0\ell} = 10^{-3} r_\ell$

Guess 2 is a damped perturbation of the true solution $\hat{G}^i$ with ε_1 representing the noise level. For convenience we shall express this level as a percentage, e.g. $\varepsilon_1 = .2$ will be expressed as $\varepsilon_1 = 20\%$ noise. The variable r_ℓ is a random complex variable with modulus uniformly distributed over $[0,1]$ and phase uniformly distributed over $[0,2\pi]$. Guess 3 represents a small totally random perturbation about the origin, again for convenience such guesses shall be denoted by $\varepsilon_1 = 100\%$.

6.2 *Choice of Step Lengths* λ_n

The first results reported are those on determination of an optimal choice of λ_n. Three possibilities were considered

1. $\lambda_n^1 = 1$
2. $\lambda_n^2 = r_n$ where r_n is a random variable uniformly distributed over $[0,2]$

3. λ_n^3 is the approximate minimum of $F(\hat{G}_n + \lambda_n \hat{H}_n)$ as a function of λ determined by Powell's cubic line search algorithm [16] with the following convergence criteria on the iterates ${}_k\lambda_n^3$

$$\left|\frac{({}_k\lambda_n^3 - {}_{k-1}\lambda_n^3)}{{}_k\lambda_n^3}\right| \cdot \left|\frac{F(\hat{G}_n + {}_k\lambda_n^3 \hat{H}_n)}{F(\hat{G}_n)}\right| < .02$$

$$\left|\frac{\nabla F(\hat{G}_n + {}_k\lambda_n^3 \hat{H}_n)}{\nabla F(\hat{G}_n)}\right| \cdot \left|\frac{F(\hat{G}_n + {}_k\lambda_n^3 \hat{H}_n)}{F(\hat{G}_n)}\right| < .02 \qquad k \leqslant 5 \tag{6.5}$$

The performances of λ_n^i were compared by running each possibility on each model problem with the appropriate test functions using the RP algorithm. The parameters $c = 2$, $N = 40$ and $N_{max} = 50$ were chosen and each problem was started with three different initial guesses with $\varepsilon_1 = 20\%$, 60% and 100%.

The results were remarkably uniform over all test cases of model problem, test function and initial guess. Choices λ_n^1 and λ_n^2 performed almost identically with λ_n^3 just under a factor of 2 better. Almost always only one extra function evaluation was required for λ_n^3, i.e. the convergence criteria of Equation (6.5) were satisfied at $k = 1$. This extra computation almost exactly balances the savings in the reduced number of iterations so that all three choices incurred the same computational cost in reduction of $F(\hat{G}_n)$ to a specified value. However the greater flexibility of λ_n^3 led to its adoption in all subsequent calculations.

The convexity of the set cI implies that $F(\hat{G}_n + \lambda \hat{H}_n) < F(\hat{G}_n)$ for all $\lambda \in (0,2)$, the computations showed that this inequality still frequently held

for $\lambda \in (2,3)$ but almost always failed beyond this. Almost all the values of λ_n^3 lay in the interval $[1,2.5]$ so in all subsequent line searches an initial guess of ${}_0\lambda_n^3 = 4$ was used. As an experiment further runs were made with the fixed choice $\lambda_n = 2$ but these compared very poorly with λ_n^1.

6.3 *Initial Guesses and Ill Conditioning*

We now present some results on the ill conditioning of the model problems. The first set are from efforts to estimate local ill conditioning through the effects of small perturbations of the data $m(v)$. They were obtained from model problems I-III with the appropriate test functions, parameters $c = 2$, $N = 40$ and $N_{max} = 40$ and using the RP algorithm. Random relative errors of size ε_2 were induced in the vector $\hat{m}$ representing the data, then the algorithm was started at the true solution $\hat{G}^i$ of the unperturbed problem. Table 1 gives some typical results for two perturbations $\varepsilon_2 = 1\%$ and $\varepsilon_2 = 5\%$. The data appearing is the percentage relative error in $\tilde{\tilde{G}}$ $(100.\|\tilde{\tilde{G}} - \hat{G}^i\|/\|\hat{G}^i\|)$ and the value $F(\tilde{\tilde{G}})$ where $\tilde{\tilde{G}}$ is the approximate solution to the perturbed problem reached by the algorithm.

The data indicate that the problems are locally well conditioned in that there exists a possible nearby solution to the perturbed problem to which the iterates tended and whose distance from the solution of the unperturbed problem is of the same order as the perturbation. We cannot assert that the solution does exist for in almost all cases the algorithm terminated with $F(G_n) < 10^{-5}$ or $n > N_{max}$. Termination due to convergence of successive

iterates did not occur although they always appeared to be circling about some common point. This problem was encountered throughout these test computations, however we delay further discussion until the presentation of results on filtering.

The second set of results indicates the global ill conditioning of the problem, particularly with respect to the natural measure $F(\hat{G})$. Figures 4-6 represent three functions $\tilde{\hat{G}}$ found as solutions to model problem I with test function $\hat{G}^2$ with values $F(\tilde{\hat{G}})$ of 8×10^{-5}, 7×10^{-5} and 8×10^{-4}, respectively. Figures 7-9 show the $\tilde{\hat{G}}$'s obtained by repeating the runs for model problem III holding all other factors (algorithm, initial guess, etc.) constant. The values $F(\tilde{\hat{G}})$ were 2×10^{-2}, 3×10^{-2} and 7×10^{-2}, respectively. Although all functions $\tilde{\hat{G}}$ resemble $\hat{G}^2$ to some degree the wide variance among them suggests that the surface $T \equiv \{(F(G),G) : G \in cI\}$ is very rugged. Furthermore although the functions in Figures 7-9 appear graphically to be closer to $\hat{G}^2$ than those in Figures 4-6 the function $F(\hat{G})$ holds them to be substantially further away.

In general the solutions $\tilde{\hat{G}}$ to model problems II and III were, as suggested by these graphs, definitely more acceptable as solutions to the phase retrieval problem than solutions to model problem I; nevertheless the values $F(\tilde{\hat{G}})$ for II and III were almost always a factor of 10 to 10^3 greater than those of I. A tentative ranking of the problems in order of decreasing ill conditioning would be I (with $\hat{G}^1$) > II > I (with $\hat{G}^2$) > III. As expected more information gives better solutions, however we were not sure to what degree the particular choice of $\hat{G}^1$ with its large discontinuity at $\omega = c$

influenced this ordering.

We conclude this subsection with a brief mention of an anomaly that lead to use of initial guess 3 rather than guess 1. As noted in a previous paper [19] it is possible to show that if iterates possess certain symmetries then almost all the algorithms preserve these symmetries. In particular with the highly symmetric choice of $\hat{G}_0 = 0$ some form of symmetry was always preserved yet the algorithm often converged to reasonable pseudo-solutions. One such example for model problem I with $N = 40$ and $c = 2$ is shown in Figure 10, it displays a symmetric solution $\tilde{\hat{G}}$ for which $F(\tilde{\hat{G}}) < 5 \times 10^{-5}$. Therefore, if no a priori knowledge is available for $\hat{G}$, a small random perturbation is a better initial guess than zero.

6.4 *Filtered PR and PA Algorithms*

Having determined the problem to be globally ill conditioned we sought to produce a better conditioned global problem by filtering out the local ill conditioning of the finite Fourier transform discussed previously. This was done by projecting the search directions $\hat{H}_n$ (and therefore the iterates $\hat{G}_n$) on to the span of the first P eigenfunctions of $P_c \mathcal{F} P_c$ whose associated eigenvalues had magnitude greater than a prescribed cutoff ε_3. As a test this filter was applied to iterates in the RP algorithm producing the filtered restricted projection algorithm (FRP) and the algorithms compared on all model problems with parameters $c = 2$, $N = 40$ and $N_{max} = 40$. Two choices of cutoff $\varepsilon_3 = .9$ and $\varepsilon_3 = .25$ were considered, these values gave projections on to subspaces of dimensions 15 and 18, respectively.

As measured by $F(G)$ FRP performed very poorly compared to RP for $\varepsilon_3 = .9$. At best 1.5 times as many iterations were needed to reduce $F(\hat{G}_n)$ to comparable values; very often FRP failed in that it ran the full N_{max} iterations reaching a point $\tilde{\hat{G}}$ such that $F(\tilde{\hat{G}}) > 100F(\overset{\times}{\hat{G}})$, where $\overset{\times}{\hat{G}}$ was the final iterate of RP under the same conditions. FRP substantially improved with $\varepsilon_3 = .25$, now averaging 1.5 times as many iterations, but still occasionally failing in the manner described above. However, if measured by the distance between final iterates the algorithms appeared to be more equal, $\|\tilde{\hat{G}} - \overset{\times}{\hat{G}}\|$ was usually less than .25 for all algorithms.

The disappointing result was that FRP did not display improved convergence properties over RP with regard to the size of $\Sigma_{k=0}^{2} \|\lambda_{n-k}\hat{H}_{n-k}\|$. Iterates still appeared to wander through the subspace without much effect on $F(G_n)$. It seems that a fifteen dimensional subspace is still too large and quite severe restrictions $(P \sim 5)$ must be used to ensure convergence to a minimum.

These results were borne out by the results and comparison of the PA and RP algorithms over all model problems for parameter pairs (ε,N) of $(2,40)$, $(2,32)$ and $(1.5,22)$ and $N_{max} = 40$. A variety of filters were used but even for the best, using the cutoff of Equation (6.6), the results had a large variance and averaged out so that PA still required about the same number of iterations to achieve the same improvement in $F(\hat{G}_n)$. Again termination was due to reduction of $F(\hat{G})$ or $n \geq N_{max}$ and not to convergence of successive iterates. One reason for this is that for the parameter values c under consideration almost all significantly nonzero eigenvalues of $P_c \mathcal{F} P_c$ have magnitude ~ 1 so that the filtered inverse $(P_c \mathcal{F} P_c)_f^{-1}$ and the Hermitian

square $P_c\mathcal{F}^{-1}P_c\mathcal{F}P_c$ agree on most functions, therefore the search directions produced by the algorithms are usually quite close. It was noticed that the relative performance of PA did appear to improve with decreasing c, averaging about .9 for $c = 1.5$.

6.5 *Relative Performance of Second Order Algorithms*

We now present results on the relative performance of algorithms FN, FNP, SQFN and SQFNP compared to RP. They are not comprehensive; considerations such as the number of parameters at the users disposal, failure of the IMSL subroutines on certain equations and computational cost prevented this. The algorithms were run on each model problem with the parameter pairs (c,N) being $(2,40)$, $(2,32)$ and $(1.5,22)$, N_{max} ranging from 15 (for computationally costly algorithms) to 50 and with perturbations in the initial guess of $\varepsilon_1 = 20$, 60 and 100. The cutoff parameters ε_3 in Equations (5.26), (5.27), (5.34) and (5.35) used at the n-th iteration, were calculated from quantity

$$\rho = \max\{\min\{4.2\|\lambda_{n-1}\hat{H}_{n-1}\| - .024, .2\}, .002\} . \tag{6.6}$$

For FN and FNP $\varepsilon_3 = \rho$, for SQFN and SQFNP $\varepsilon_3 = \rho/3$. The IMSL subroutine EIGRS was used to calculate the eigendecomposition of $\nabla^2 F(\hat{G}_n)$, the subroutine LLSQF to calculate the minimum norm least squares solution of Equation (5.42), with the upper bound ATOL on the condition number of R in the QR factorization specified to be $\varepsilon_3/10$.

The results in Tables 2-5 give first the average performance of RP on each problem in terms of the final value $F(\tilde{\hat{G}})$ and the number of iterates $\tilde{n}$ needed to reach this. The remaining entries are the ratios $\check{n}/\tilde{n}$ where $\check{n}$ is the average number of iterations needed by the alternative algorithm to reduce $F(\hat{G}_{\check{n}})$ to below $F(\tilde{\hat{G}})$. In some cases the algorithm on trial failed to reduce $F(\hat{G}_{\check{n}})$ to below $100F(\tilde{\hat{G}})$, such cases are marked by a *. In other cases the alternative algorithm converged but with $F(\hat{G}_{\check{n}})$ larger than $F(\hat{G}_{\tilde{n}})$ in which case the $\tilde{n}$ of the ratio was altered to the number of iterations required by RP to reduce $F(\hat{G}_{\tilde{n}})$ to below $F(\hat{G}_{\check{n}})$.

The remaining tables show the effect of (and thus the need for) filtering. Table 6 contains pairs giving the range of the number of eigenvalues preserved at each iteration by the filter in FN and FNP. The final table for SQFN and SQFNP shows two pairs, the upper giving the range of the number of rows in the filtered block equation (5.40), the lower pair indicates the range of the number of columns used by LLSQF. Typically the larger figures in the range occur as $F(\hat{G}_n)$ decreases and $\hat{G}_n$ converges, however the "stop-start" behaviour of the iterates often caused these measures of filtering to also jump about.

Finally we present graphs (Figures 11-22) of final iterates for RP with parameters $c = 2$ and $N = 40$ for each model problem and $\varepsilon_1 = 20\%$, 60% and 100%. As noted previously final iterates varied much less than final values of $F(\hat{G}_n)$ and the graphs are typical of all final iterates for these problems.

Some points worth noting on observed algorithm behaviour are given. RP hardly ever halted due to convergence of the sequence $\hat{G}_n$, iterates exhibited

steady "zig-zagging" with accompanying small but steady decreases in $F(\hat{G}_n)$ after an initial burst of reduction in the first 10 or so iterates. The higher order algorithms did give convergent iterates on a number of occasions but some limit points were not true minima since at $\hat{G}_n$ $\nabla F(\hat{G}_n) \neq 0$ but $\nabla F(\hat{G}_n)$ lay in the null space of the filtered Hessian.

The cutoff parameter given by Equation (6.6) was derived after some trials and errors and is certainly problem specific. However it performed satisfactorily for FN, FNP, SQFN and SQFNP in that it generated λ_n that almost always lay in the interval [.1,1.5], and for FN and FNP started close to the true solution the expected value of $\lambda_n \sim 1$ was always observed. Moreover it was noted that RP gave the best initial decreases in $F(\hat{G}_n)$ when started with $\varepsilon_1 = 100\%$, so in all algorithms the final search direction $\hat{H}_n$ was taken to be a weighted linear combination of the $\hat{H}_n^1$ determined by the particular algorithm and the $\hat{H}_n^2$ of RP, i.e.

$$\hat{H}_n = \hat{H}_n^1 \left(\frac{.1}{.1+\varepsilon_3}\right) + \hat{H}_n^2 \left(\frac{\varepsilon_3}{.1+\varepsilon_3}\right) \quad . \tag{6.7}$$

7. CONCLUSIONS

The ill posed nature of phase retrieval induces enough variance in the data to prevent the drawing of quantitative conclusions, however qualitative results are apparent. As expected FN and FNP show the best performance when iterates are close to the true solution $\hat{G}^i$, with FNP displaying greater robustness in regions further away. Close to the solution SQFN and SQFNP do not improve upon RP, but they give the best results when started from a random initial guess. Since SQFN, SQFNP, FN and FNP are all filtered variants of the same basic algorithm it should be possible by a judicious choice of filters to combine the best features of all in one algorithm; to do so requires further study of the block filterings of subsection 5.5.

As measured in terms of the computational cost to reduce $F(\hat{G}_n)$ to a certain value RP appears to be the certain winner and it is difficult to see how any other algorithm can be improved enough to compete with it. Of the second order algorithms SQFNP seems to be the most efficient, requiring a matrix with .8 the number of rows as SQFN to give approximately the same results. Furthermore the cost can probably be substantially reduced by taking advantage of the sparsity and special structure of the blocks in Equation (5.42).

Tables 6 and 7 confirm that the essential dimension of model problem I is twice the number P of significantly nonzero eigenvalues of $P_c \mathcal{F} P_c$. The essential dimension of II and III is harder to estimate although a reasonable upper bound would be $2P + N$. Certainly for $N = 22$ and 32 the problems

appear to be almost well posed. The tables also indicate that II may be more (locally) ill conditioned than III although this may be attributable to the form of the test functions.

The oscillatory nature of the graphed solutions suggests that further filtering is necessary and that local filtering to the degree done here is insufficient to give a well posed problem. Three options for further conditioning are:

1. Use of a more restrictive local filter
2. Regularization of $F(G)$ by addition of a penalty function e.g.,

$$F_R(G) = F(G) + \alpha \left\| \frac{dG}{d\omega} \right\|^2 \tag{7.1}$$

3. Use of the algorithm above followed by smoothing of the numerical solutions $\hat{G}^*$.

Each option has its attractions: with use of 1 the iterates would be restricted to the span of the first 5 to 10 singular functions, these are known to be smooth and slowly varying so a linear combination would also be reasonably well behaved. Note that this also appears if the interval size c is reduced to around .75 or smaller, thus having less information may actually give a better solution. For 2 a suitably chosen penalty function can easily be accommodated within the theory of algorithms based on projections developed here. Finally visual inspection of the graphs shown indicates that the functions $\hat{G}^*$ are often very perturbed but are basically similar to the true solution $\hat{G}^i$ so that option 3 will often give an acceptable final solution.

APPENDIX A. ON THE EXISTENCE OF MULTIPLE SOLUTIONS TO THE 2-D PHASE RECOVERY PROBLEM

The two-dimensional phase retrieval problem can be stated as:

Let A and B be bounded subsets of $\mathbb{R}^2$. Given the information that $g(z_1,z_2)$ is the Fourier transform of a function $G(\omega_1,\omega_2)$ with support contained in B and the values $m(x_1,x_2) = |g(x_1,x_2)|$ on A, find the phase of g on A and reconstruct G.

The phase retrieval problem does not necessarily have a unique solution. The aim of this appendix is to derive from a given solution pair g and G *necessary* conditions for the existence (and characterization) of alternative solution pairs h and H. An intuitive start for such an investigation is the simple result that if $f(z)$ is an analytic function of the complex variable z then so is $f^*(z^*)$ and $f(z)$ and $f^*(z^*)$ have the same modulus for real z. This suggests that if a solution g can be factored into a product of analytic functions g_1 and g_2 then the entire function of two complex variables $h(z_1,z_2) \equiv g_1(z_1,z_2)g_2^*(z_1^*,z_2^*)$ is also a possible solution. The main result of this appendix is to show that all possible alternative solutions must be of this simple form.

A.1 *One-Dimensional Results*

The following results from the theory of functions of a single complex variable are required.

Theorem A.1: Paley-Wiener Theorem [20]. Let $B \equiv [b_1,b_2]$ be a bounded interval in $\mathbb{R}$. Then for any $G \in L^2(\mathbb{R})$, $G \neq 0$ on B, the transform

$$g(z) = \int_{b_1}^{b_2} e^{izu} G(u)\, du \tag{A.1}$$

is an entire function and there exist constants α and β such that

$$|g(z)| < \begin{cases} \alpha\, e^{-b_1 \operatorname{Im} z} & \text{if } \operatorname{Im} z > 0 \\ \beta\, e^{b_2 \operatorname{Im} z} & \text{if } \operatorname{Im} z < 0 \end{cases} . \tag{A.2}$$

The next result is the fundamental theorem providing the necessary machinery to characterize all possible solutions to the phase problem both in one and two dimensions. Although independently derived by many authors [3,21], it appears to have been first stated by Akutowicz [8.9]. We state the result as originally presented there.

Theorem A.2. Let $\mathscr{C}$ be the class of all functions $g \in L^2(\mathbb{R})$ satisfying

1. $|g(x)| = m(x) \neq 0 \qquad \forall x \in \mathbb{R}$
2. $g = \mathscr{F}G$ where support of G is contained in a bounded interval B of $\mathbb{R}$.

Then any two functions $g, h \in \mathscr{C}$ are related by equations of the form

$$\begin{aligned} h(z) &= e^{i(\alpha+\beta z)} B(z) g(z) \\ B(z) &= \prod_{\ell=1}^{\infty} \left(\frac{z - z_\ell^*}{z - z_\ell} \right) \end{aligned} \tag{A.3}$$

where the z_ℓ form some subset of the zeros of $g(z)$. The function $(z-z_\ell^*)/(z-z_\ell)$ is termed a Blaschke factor.

Lemma A.1: A necessary and sufficient condition for the infinite product $B(z)$ to converge is that [9]

$$\sum_{\ell=1}^{\infty} \frac{|\text{Im } z_\ell|}{1+|z_\ell|^2} < \infty \quad . \tag{A.4}$$

A sufficient condition for the convergence of the infinite sum is that $G(u)$ have only a finite number of jump discontinuities over B.

It is easily shown that if G has support in an interval B and if $h(z) = B(z)g(z)$, then $H(u) = (\mathcal{F}^{-1}h)(u)$ also has support in B. Therefore combining theorems A.1, A.2 and lemma A.1 gives the following statement on existence of multiple solutions to the 1-D phase retrieval problem.

Theorem A.3: Let A and B be bounded intervals in $\mathbb{R}$ with a modulus $m(x)$ specified over A and a solution pair g and G be given to the corresponding 1-D phase problem. Then if $m(x)$ has an extension to the entire real line such that $m(x) > 0$ and $G(u)$ has only finite jump discontinuities over B as well as being nonzero in neighborhoods of the endpoints of B, then all other solution pairs h,H are given by

$$h(z) = e^{i\alpha}B(z)g(z) \tag{A.5}$$

$$H(u) = (\mathcal{F}^{-1}h)(u) \tag{A.6}$$

where $B(z)$ is any finite or infinite product of Blaschke factors and $\alpha \in \mathbb{R}$.

The conditions of Theorem A.3 imply that a solution $g(z)$ has a Hadamard factorization

$$g(z) = |g(0)|e^{i(\alpha+\beta z)} \prod_{\ell=1}^{\infty} \left(1 - \frac{z}{z_\ell}\right) \tag{A.7}$$

where $\alpha,\beta \in \mathbb{R}$, which may be rewritten as

$$g(z) = \left[|g(0)|e^{i(\alpha+\beta z)} \prod_{\ell\in\Lambda} \left(1 - \frac{z}{z_\ell}\right)\right]\left[\prod_{\ell\in(N-\Lambda)} \left(1 - \frac{z}{z_\ell}\right)\right]$$

$$\equiv g_1(z)g_2(z) \tag{A.8}$$

where Λ is a subset of the natural numbers N. Theorem A.3 thus states that any solution $h(z)$ has the form

$$h(z) = e^{i\gamma} g_1(z)g_2^*(z^*), \qquad \gamma \in \mathbb{R} \tag{A.9}$$

i.e. that all possible solutions are in one to one correspondence with all possible factorizations of $g(z)$.

A.2 *Extension to Two Dimensions*

Conditions that multiple solutions to the 2-D phase retrieval problem must satisfy can be deduced from the 1-D results. To begin, suppose that the problem as stated has a solution pair $g(z_1,z_2)$, $G(\omega_1,\omega_2)$ where $z = x + iy$ and $\omega = u + iv$ denote variables in the transform and physical domains, such that $G(u_1,u_2)$ has only a finite number of jump discontinuities over B. Then

Lemma A.2: $g(z_1,z_2)$ is an entire function of the complex variables z_1,z_2 of exponential growth.

Proof. After defining the quantities

$$u_1^+ = \max\{u_1 : (u_1,u_2) \in B\}, \quad u_2^+(u_1) = \max\{u_2 : (u_1,u_2) \in B\}$$

$$u_1^- = \min\{u_1 : (u_1,u_2) \in B\}, \quad u_2^-(u_1) = \min\{u_2 : (u_1,u_2) \in B\}$$

$g(z_1,z_2)$ can be expressed as

$$g(z_1,z_2) = \int_{u_1^-}^{u_1^+} du_1\, e^{iz_1u_1} \int_{u_2^-(u_1)}^{u_2^+(u_1)} du_2\, e^{iz_2u_2} G(u_1,u_2) . \quad (A.10)$$

Writing $g(z_1,z_2)$ as $g_{z_2}(z_1)$ to indicate that $g(z_1,z_2)$ is to be considered as a function of z_1 only with z_2 fixed gives

$$g_{z_2}(z_1) = \int_{u_1^-}^{u_1^+} \tilde{G}(u_1,z_2)e^{iz_1u_1} du_1 \quad (A.11)$$

Therefore by Theorem A.1 $g_{z_2}(z_1)$ is an entire function of z_1 of exponential growth. A similar procedure shows that $g_{z_1}(z_2)$ is an entire function of z_2.

An immediate consequence of this lemma is that if a solution exists then the modulus $m(x_1,x_2)$ has an analytic extension to all of $\mathbb{R}^2$, which, under the assumption that $m(x_1,x_2) > 0 \ \forall (x_1,x_2)$, is unique.

Now let $h(z_1,z_2)$, $H(\omega_1,\omega_2)$ be any other solution pair to this problem; then $h_{x_2}(x_1)$ and $g_{x_2}(x_1)$ must have the same modulus $m_{x_2}(x_1)$ over the set $B_{x_2} = \{x_1 : (x_1,x_2) \in B\}$; i.e. $h_{x_2}(z_1)$ and $g_{x_2}(z_1)$ are both solutions to a 1-D phase retrieval problem. Therefore by Theorem A.2, Lemma A.1 and the above assumptions on g, G and m we have that

$$h_{x_2}(z_1) = e^{i\alpha(x_2)} e^{i\beta(x_2)z_1} B_{x_2}(z_1) g_{x_2}(z_1) \tag{A.12}$$

where $\alpha(x_2)$ and $\beta(x_2)$ are constants dependent on x_2 only and $B(z)$ is an infinite product of Blashke factors formed from the zeros $\rho_\ell(x_2)$ of $g_{x_2}(z_1)$. Thus if $\eta_\ell(x_2)$ are the zeros of $h_{x_2}(z_1)$ and the sets X_{x_2}, Y_{x_2} are defined by

$$X_{x_2} \equiv \bigcup_{\ell=1}^{\infty} \{\rho_\ell(x_2), x_2\}, \qquad Y_{x_2} \equiv \bigcup_{\ell=1}^{\infty} \{\eta_\ell(x_2), x_2\} \tag{A.13}$$

it follows that

$$Y_{x_2} \subseteq X_{x_2} \cup X^*_{x_2}, \qquad X_{x_2} \subseteq Y_{x_2} \cup Y^*_{x_2} . \tag{A.14}$$

Let X and Y be the sets of zeros of g and h respectively. It is known [22] that the zeros of a function of n complex variables form an analytic set of dimension $(n-1)$ which in turn is the union of analytic manifolds of dimension $(n-1)$ and a set of singular points of dimension at most $(n-2)$. Therefore for almost all x_2 the points $(\rho_k(x_2), x_2)$ and $(\eta_k(x_2), x_2)$ are members of analytic submanifolds of X and Y respectively; that is for each k there exist maps

$$\begin{aligned} \phi_k &: C^1 \to X & \phi_k(x_2) &= (\rho_k(x_2), x_2) \\ \psi_k &: C^1 \to Y & \psi_k(x_2) &= (\eta_k(x_2), x_2) \end{aligned} \tag{A.15}$$

such that for almost all x_2, ϕ_k and ψ_k are analytic in a neighborhood $N_k(x_2)$ of x_2. Furthermore for almost all x_2, the maps

$\{\phi_k, \psi_k\}_{k=1}^{\infty}$ are analytic in neighborhoods $N_k(x_2)$ of x_2 (note the dependence of $N_k(x_2)$ on k).

We now suppose that the zeros $\rho_k(x_2)$ are well separated, i.e.

$$\rho_k(x_2) \neq \rho_\ell(x_2) \quad \text{or} \quad \rho_\ell^*(x_2) \qquad \forall k,\ell \quad k \neq \ell \tag{A.16}$$

Since by (A.12) $y_k(x_2) = \rho_k(x_2)$ or $\rho_k^*(x_2)$ the analyticity of ϕ_k and ψ_k imply that

$$y_k(z) = \rho_k(z) \quad \text{or} \quad \rho_k^*(z^*) \qquad \forall z \in N_k(x_2) \ . \tag{A.17}$$

Therefore if $X_1 \subseteq X$ and $Y_1 \subseteq Y$ are the analytic extensions of the neighborhoods of X_{x_2} and Y_{x_2} to analytic manifolds then

$$X_1 \subseteq Y_1 \cup Y_1^* \ , \qquad Y_1 \subseteq X_1 \cup X_1^* \ . \tag{A.18}$$

If $Y_1 \neq X_1$ *then there exists a point* $y \in Y_1$ *and an associated neighborhood* $N(y) \subset Y_1$ such that $N(y) \subset X_1^* - X_1$. By analytic continuation $N(y)$ can be extended to an analytic manifold Y_2 such that $Y_2 \subseteq X_1^*$ and $Y_2 \subseteq Y_1$. If $Y_2 = Y_1$ then $Y_1 = X_1^*$ otherwise Y_1 must be decomposable into two analytic submanifolds Y_2 and Y_3 such that $Y_2 \subset X_1^*$ and $Y_3 \subset X_1^*$ or X_1.

Thus, apart from alternative solutions generated by varying $\alpha(x_2)$ and $\beta(x_2)$, a necessary condition that an alternative solution h to the 2-D phase problem must satisfy is that some of its zeros be the complex conjugates of those of the original solution. Although this is the same mechanism by which an infinite number of alternative solutions to the 1-D phase problem are generated, the zeros must now satisfy the condition that they form a union of one-dimensional analytic manifolds as opposed to a union of zero

dimensional manifolds, that is a collection of connected analytic line segments as opposed to a collection of isolated points. If an alternative solution exists then either the whole manifold X has been "flipped" to its conjugate, or it has been "torn" and only partially flipped. The connected nature of X_1 implies that the existence of "dotted lines" along which tears may be made is very unlikely; this compares to the isolated points in the 1-D problem, each of which may be flipped independently of the other.

Given this condition the form of an alternative solution may be determined. Let X_1 be decomposable into submanifolds X_2, X_3 and define

$$X_{i,x_2} \equiv X_i \cap \{(z_1,x_2): x_2 \text{ fixed}, z_1 \in C^1\} \quad . \tag{A.19}$$

then the function $g_{x_2}(z_1)$ may be written as the product

$$g_{x_2}(z_1) = g_{1,x_2}(z_1) g_{2,x_2}(z_1) \tag{A.20}$$

where

$$g_{2,x_2}(z_1) = \prod_{(\rho_k(x_2),x_2) \in X_{2,x_2}} \left(1 - \frac{z_1}{\rho_k(x_2)}\right) \quad . \tag{A.21}$$

By (A.12) $h_{x_2}(z_1)$ may be written as

$$h_{x_2}(z_1) = g_{1,x_2}(z_1) g^*_{2,x_2}(z_1^*) \quad . \tag{A.22}$$

If $h(z_1,z_2)$ exists it is the analytic extension of $h_{x_2}(z_1)$, therefore $h(z_1,z_2) = g_1(z_1,z_2) g_2^*(z_1^*,z_2^*)$.

We have not been able to show that this necessary condition for alternative solutions is also sufficient; i.e. given submanifolds X_2, X_3 and

the decomposition of Eq. (A.20) that the $h_{x_2}(z_1)$ of Eq. (A.22) may be analytically continued to a function $h(z_1,z_2)$. One source of trouble is the dependence of $N_k(x_2)$ on k; it is possible that for every x_2 $\cap_{k=1}^{\infty} N_k(x_2) = \phi$. Then although each zero $(\rho_k(x_2),x_2)$ is analytic in a neighborhood of x_2 there does not exist a neighborhood over which all zeros are uniformly analytic, and therefore a neighborhood over which the product of Eq. (A.21) is provably analytic.

If the cardinalities of X_{i,x_2} are finite, e.g. $g(z_1,z_2)$ is a polynomial,.then sufficiency can be shown. In the polynomial case decomposability of X_1 into X_2 and X_3 is equivalent to a factorization of the polynomial. However almost all polynomials of two variables are irreducible so that such a factorization and decomposition does not exist, therefore alternative solutions do not exist. Irreducibility extends to general functions of two variables with infinite sets of zeros, so that exact alternative solutions are most unlikely in 2-D phase retrieval. This result on polynomials and its implications is also presented in [23].

A.3 *The Support of Alternative Solutions*

In the previous subsection conditions that alternative solutions g and h must satisfy in order that $|g(x_1,x_2)| = |h(x_1,x_2)|$ were derived. It remains to derive necessary conditions on g and h so that (support G) = (support H). The first is that $\beta(x_2) \equiv 0$ in Eq. (A.12). This follows by noting that if $G(u_1,u_2)$ is nonzero in neighborhoods of points (u_1^+,u_2), $(u_1^-,u_2) \in B$ then the function $\tilde{G}(u_1,x_2)$ of Eq. (A.11) will be nonzero in

neighborhoods of $u_1 = u_1^+$ and $u_1 = u_1^-$ for almost all x_2. So by Theorem A.3 $h_{x_2}(z_1)$ is the transform of a function $\tilde{H}(u_1,x_2)$ with support in (u_1^-,u_1^+) if and only if $\beta(x_2) \equiv 0$.

A second condition follows from noting that the boundedness of the set B implies that $h(z_1,z_2)$ is of exponential growth in z_2, so that $\alpha(x_2)$ must only be a linear function of x_2. Summarizing these results and those of the previous section gives the next theorem.

Theorem A.4: Let g, G be a solution pair to the 2-D phase retrieval problem. Then any other solution pair h, H must have the form

$$h(z_1,z_2) = e^{i(\alpha_1+\alpha_2 z_2)} g_1(z_1,z_2) g_2^*(z_1^*,z_2^*) \tag{A.23}$$

where $g_1 g_2$ is a factorization of g.

We have been unable to complement these necessary conditions for equality of support with sufficient conditions equivalent to those for the 1-D problem. The difficulty seems to lie in determining the role of the geometry of B; we give two examples

1. The first example concerns convexity and is taken from Huiser and Torn [24]. Let g, G be a solution pair, then after the change of variables to the new orthogonal coordinate systems $(s_1,s_2),(t_1,t_2)$ with

$$\begin{aligned} s_1 &= u_1 \cos\psi + u_2 \sin\psi & t_1 &= x_1 \cos\psi + x_2 \sin\psi \\ s_2 &= -u_1 \sin\psi + u_2 \cos\psi & t_2 &= -x_1 \sin\psi + x_2 \cos\psi \end{aligned} \tag{A.24}$$

and definition of the quantities

$$s_2^+(s_1) = \max\{s_2 : (s_1,s_2) \in B\}$$

$$s_2^-(s_1) = \min\{s_2 : (s_1,s_2) \in B\}$$

$$s_1^+(\psi) = \max\{s_1 : (s_1,s_2) \in B\} \tag{A.25}$$

$$s_1^-(\psi) = \min\{s_1 : (s_1,s_2) \in B\}$$

the relationship $g = \mathcal{F}G$ may be rewritten as

$$g(t_1,t_2) = \int_{s_1^-(\psi)}^{s_1^+(\psi)} ds_1\, e^{is_1t_1} \cdot \int_{s_2^-(s_1)}^{s_2^+(s_1)} ds_2\, e^{is_2t_2}\, G(s_1,s_2) \quad . \tag{A.26}$$

For fixed t_2 the growth rate in $g_{t_2}(t_1)$ is determined by $s_1^+(\psi)$ and $s_1^-(\psi)$. Knowing these values for all ψ is equivalent to knowing all *supporting hyperplanes for the set B, which by duality arguments from linear* algebra is equivalent to knowing the convex hull of B. If h and H is any other solution pair then $h_{t_2}(t_1)$ must have the same growth as $g_{t_2}(t_1)$ otherwise H has support outside of the convex hull of B.

If g has a factorization g_1g_2 such that the growth of g_2 is always dominated by that of g_1 (e.g. g_2 is a polynomial) then the alternative function

$$h(z_1,z_2) = g_1(z_1,z_2)g_2^*(z_1^*,z_2^*) \tag{A.27}$$

has the same modulus as g and support in the convex hull of B. If B is convex then h is an alternative solution, if B is not convex then it is possible that the support of H is not B even though still in the convex hull.

2. Let g, G be a solution pair, then it is trivial to show that the inverse transform of $g^*(z_1^*,z_2^*)$ is $G^*(-\omega_1,-\omega_2)$ which has support $-B$. So a sufficient condition that $g^*(z_1^*,z_2^*)$ be an alternative solution is that $B = -B$, i.e. B is invariant under rotation by 180°.

Example 1 suggests that convexity of B is necessary for existence of an alternative solution and taken with example 2 suggests that for a factorization g into $g_1 g_2$ and an alternative solution h of Eq. (A.27) then B must have symmetries linked in some fashion to those directions in which growth of g_2 dominates g_1.

A.4 *Conclusions*

Nonuniqueness in the phase retrieval problem in two dimensions appears to depend on two conditions: (1) that the zero space of g be decomposable into a union of several submanifolds, (2) that B possesses a suitable combination of convexity and symmetry. Both conditions will, in general, be difficult to satisfy compared to the 1-D phase retrieval problem. Only in the case of symmetries that effectively reduce $g(z_1,z_2)$ to a function of one variable (e.g., the possession of radial symmetry investigated in [25]) will the manifold have an infinite decomposition as appears in the 1-D problem. In most cases it will be indivisible. Likewise the general two-dimensional bounded set has considerably more degrees of freedom than the one-dimensional bounded set, the interval, consequently it has far fewer symmetries. Therefore in general the 2-D phase retrieval problem will have a unique solution if one exists.

REFERENCES

1. R. Barakat and G. Newsam, "Algorithms for reconstruction of partially known, bandlimited Fourier transform pairs from noisy data: I the prototypical linear problem." Submitted for publication to J. Integral Eqs.
2. D.L. Misell, "The phase problem in electron microscopy," in Advances in Optics and Electron Microscopy, Vol. 7, eds. V.E. Coslett and R. Barer (Academic Press, New York, 1978).
3. R. Burge, M. Fiddy, A. Greenway, and G. Ross, "The phase problem," Proc. Roy. Soc., A350, 191-212 (1976).
4. J.R. Fienup, "Phase retrieval algorithms: a comparison," Appl. Opt., 21, 2758-2769 (1982). See other works by the author referenced here.
5. B.R. Frieden, "Image enhancement and restoration," in Picture Processing and Digital Filtering, 2nd ed., ed. T.S. Huang (Springer-Verlag, New York, 1979).
6. H.A. Ferwerda, "The phase reconstruction problem for wave amplitudes and coherence functions," in Inverse Source Problems in Optics, ed. H.P. Baltes (Springer-Verlag, New York, 1978).
7. W. Saxton, Computer Techniques for Image Processing in Electron Microscopy (Academic Press, New York, 1978).
8. E.J. Akutowicz, "On the determination of the phase of a Fourier integral, I." Trans. Amer. Math. Soc., 83, 179-192 (1956).
9. E.J. Akutowicz, "On the determination of the phase of a Fourier integral, II," Proc. Amer. Math. Soc., 8, 234-238 (1957).

10. A.N. Tikhonov and V.Y. Arsenin, Solutions of Ill-Posed Problems (Halsted, New York, 1977).

11. J.N. Chapman, "The application of iterative techniques to the investigation of strong phase objects in the electron microscope," Phil. Mag., 32, 527-552 (1975).

12. L. Gubin, B. Polyak, and E. Raik, "The method of projections for finding the common point of convex sets," USSR Comp. Math. and Math. Phys., 7, 1-24 (1967).

13. R. Gerschberg and W. Saxton, "A practical algorithm for the determination of phase from image and diffraction plane pictures," Optik, 35, 237-246 (1972).

14. M.S. Sabri and W. Steenart, "An approach to bandlimited extrapolation: the extrapolation matrix," IEEE Trans. Circuits Syst., CAS25, 74-78 (1978).

15. M.J.D. Powell, "An efficient method for finding the minimum of a function of several variables without calculating derivatives," Compt. J., 7, 155-164 (1964).

16. G. Walsh, Methods of Optimization (Wiley, New York, 1975).

17. J.E. Dennis, Jr., "Nonlinear least squares and equations," in The State of the Art in Numerical Analysis, ed. D. Jacobs (Academic Press, New York, 1977).

18. R.B. Holmes, "Smoothness of certain metric projections on Hilbert space," Trans. Am. Math. Soc., 183, 87-100 (1973).

19. R. Barakat and G. Newsam, "A numerically stable iterative method for the inversion of wavefront aberrations from measured point spread function data," J. Opt. Soc. Am., 70, 1255-1263 (1980).

20. P.P. Boas, Entire Functions (Academic Press, New York, 1954).

21. L.S. Taylor, "The phase retrieval problem," IEEE Trans. Antennas Prop. AP39, 381-391 (1981).

22. L. Ronkin, Introduction to the Theory of Entire Functions of Several Complex Variables (Amer. Math. Soc., Providence, 1974).

23. Yu. M. Bruck and L.G. Sodin, "On the ambiguity of the image reconstruction problem," Opt. Comm., 30, 304-308 (1979).

24. A. Huser and P. van Torn, "Ambiguity of the phase reconstruction problem," Opt. Lett., 5, 499-501 (1980).

25. W. Lawton, "Uniqueness results for the phase retrieval problem for radial functions," J. Opt. Soc. Am., 71, 1519-1522 (1981).

Table 1. Sensitivity of solution $\hat{\tilde{G}}$ to a relative perturbation of ε_2 in the data.

		$\varepsilon_2 = 1\%$		$\varepsilon_2 = 5\%$	
Model Problem	Test Function	Percentage error in $\hat{\tilde{G}}$	$F(\hat{\tilde{G}})$	Percentage error in $\hat{\tilde{G}}$	$F(\hat{\tilde{G}})$
I	$\hat{G}^1$	1%	1×10^{-5}	9%	6×10^{-4}
I	$\hat{G}^2$	2%	9×10^{-6}	10%	3×10^{-5}
II	$\hat{G}^1$	.75%	8×10^{-4}	4%	4×10^{-2}
III	$\hat{G}^2$	1.5%	2×10^{-4}	13%	3×10^{-3}

Table 2. Relative performance of algorithms on model problem I with test function G^1.

ε_1	N	RP		FN	FNP	SQFN	SQFNP
	22	8×10^{-6}	6	.325	.325	1.6	1.8
20	32	9×10^{-6}	14	.275	.275	.775	1.3
	40	5×10^{-6}	5	.80	.80	2.0	2.0
	22	9×10^{-6}	34	.20	.10	.275	.40
60	32	1×10^{-5}	7	.575	.575	2.0	1.4
	40	9×10^{-6}	10	.40	.40	1.5	1.5
	22	5×10^{-4}	50	*	.65	.45	.50
100	32	7×10^{-4}	30	.65	.45	.40	.40
	40	1.4×10^{-3}	40	*	.20	.30	.30

Table 3. Relative performance of algorithms on model problem I with test function G^2.

ε_1	N			FN	FNP	SQFN	SQFNP
	22	1.9×10^{-5}	50	1.5	1.0	.40	.30
60	32	1.5×10^{-5}	20	*	.825	.875	.45
	40	3×10^{-5}	40	*	1.0	.35	.30
	22	5×10^{-5}	50	1.5	.30	.60	.30
100	32	4×10^{-4}	20	1.6	.45	.35	.40
	40	8×10^{-4}	40	.50	.50	.425	.30

Table 4. Relative performance of algorithms on model problem II.

ε_1	N			FN	FNP	SQFN	SQFNP
20	40	9×10^{-6}	9	.50	.50	-	1.0
60	40	9×10^{-6}	11	.50	.50	-	-
100	40	1×10^{-3}	40	.25	.20	-	-

Table 5. Relative performance of algorithms on model problem III.

ε_1	N			FN	FNP	SQFN	SQFNP
20	22	9×10^{-6}	23	.20	.20	.50	.50
60	22	5×10^{-4}	50	.50	.30	.35	.45
	32	3×10^{-4}	20	2.0	.45	.50	.60
100	22	2×10^{-3}	50	3.0	.80	.325	.325
	32	6×10^{-2}	20	*	1.0	.50	.40

Table 6. The range of nonzero eigenvalues in algorithms FN and FNP. Note that for $c = 1.5, 2$ the finite Fourier transform has at most 15 and 22 eigenvalues respectively with magnitudes greater than .002.

	N	22	22	22	32	32	40	40
	ϵ_1	20	60	100	60	100	60	100
G^1	FN	18-19	18-25	14-28	30	24-39	30-34	25-37
G^1	FNP	16-18	17-18	14-21	28	23-28	30-33	23-37
Model Problem I								
G^2	FN	18	16-27	16-28	26-43	27-40	28-39	28-35
G^2	FNP	17	14-24	16-23	24-29	25-30	24-26	28-29
Model Problem II								
	FN	-	-	-	-	-	55-57	54-60
	FNP	30-31	30-33	-	-	-	55-57	55-62
Model Problem III								
	FN	33-44	33-44	33-44	47-64	57-63	-	-
	FNP	31-37	31-42	31-42	46-61	52-58	55-75	-

Table 7. The range of matrix dimensions and effective dimensions of the block matrices in algorithms SQFN and SQFNP. The rows marked A contain the range of block matrix row sizes; the rows marked B contain the range of the number of columns used by the routine LLSQF.

		N	22	22	22	32	32	40	40
		ε_1	20	60	100	60	100	60	100
G^1	SQFN	A	40-44	30-44	32-44	36-64	44-64	49-80	54-80
		B	12-23	12-17	9-22	9-25	12-26	12-29	13-29
G^1	SQFNP	A	23-32	28-36	27-37	36-50	38-51	46-61	47-62
		B	10-18	11-17	9-24	14-20	13-28	12-26	15-32
MODEL PROBLEM I									
G^2	SQFN	A	28-44	36-44	32-44	46-64	53-63	72-80	75-80
		B	10-22	13-21	10-23	8-27	12-25	16-33	23-33
G^2	SQFNP	A	26-29	30-35	29-36	45-53	41-51	54-60	58-65
		B	14-19	10-16	10-20	18-24	11-28	11-27	20-30
MODEL PROBLEM II									
	SQFN	A	87-88	79-88	72-88	91-128	120-126	-	-
		B	36-44	25-44	18-44	31-64	33-43	-	-
	SQFNP	A	53-65	58-77	61-73	85-104	102-107	-	-
		B	23-37	17-44	18-44	31-64	30-41	-	-

FIGURE LEGENDS

Fig. 1. Two examples of poor convergence of the alternating projection algorithm induced by ill conditioning.

Fig. 2. The tangent plane $T_i(x)$ to a set B_i at the point $P_i(x)$. $T_i(x)$ locally separates B_i and $S_i(x)$.

Fig. 3. An example of the form of the Hessian operator $\mathscr{H}_i(x)$ associated with a sphere B_i.

Fig. 4. A solution to model problem I, test function $\tilde{\hat{G}}^2$ with $F(\tilde{\hat{G}}) = 8 \cdot 10^{-5}$: Δ reconstructed modulus, o reconstructed phase, ---- exact modulus, —— exact phase.

Fig. 5. A solution to model problem I, test function $\tilde{\hat{G}}^2$ with $F(\tilde{\hat{G}}) = 7 \cdot 10^{-5}$: Δ reconstructed modulus, o reconstructed phase, ---- exact modulus, —— exact phase.

Fig. 6. A solution to model problem I, test function $\tilde{\hat{G}}^2$ with $F(\tilde{\hat{G}}) = 8 \cdot 10^{-4}$: Δ reconstructed modulus, o reconstructed phase, ---- exact modulus, —— exact phase.

Fig. 7. A solution to model problem III, test function $\tilde{\hat{G}}^2$ with $F(\tilde{\hat{G}}) = 2 \cdot 10^{-2}$: Δ reconstructed modulus, o reconstructed phase, ---- exact modulus, —— exact phase.

Fig. 8. A solution to model problem III, test function $\tilde{\hat{G}}^2$ with $F(\tilde{\hat{G}}) = 3 \cdot 10^{-2}$: Δ reconstructed modulus, o reconstructed phase, ---- exact modulus, —— exact phase.

Fig. 9. A solution to model problem III, test function $\tilde{\hat{G}}^2$ with $F(\tilde{\hat{G}}) = 7 \cdot 10^{-2}$: Δ reconstructed modulus, o reconstructed phase, ---- exact modulus, —— exact phase.

Fig. 10. An example of preservation of symmetry, a solution to model problem I, test function $\tilde{\hat{G}}^1$ with $\hat{G}_0 = 0$.

Fig. 11. Sample realization solution to model problem I, test function $\tilde{\hat{G}}^1$, $\varepsilon = 20\%$: Δ reconstructed real component, o reconstructed imaginary component, —— exact real component, ---- exact imaginary component.

Fig. 12. Sample realization solution to model problem I, test function $\tilde{\hat{G}}^1$, $\varepsilon = 60\%$: Δ reconstructed real component, o reconstructed imaginary component, —— exact real component, ---- exact imaginary component.

Fig. 13. Sample realization solution to model problem I, test function $\tilde{\hat{G}}^1$, $\varepsilon = 100\%$: Δ reconstructed real component, o reconstructed imaginary component, —— exact real component, ---- exact imaginary component.

Fig. 14. Sample realization solution to model problem III, test function $\tilde{\hat{G}}^1$, $\varepsilon = 20\%$: Δ reconstructed real component, o reconstructed imaginary component, —— exact real component, ---- exact imaginary component.

Fig. 15. Sample realization solution to model problem III, test function $\tilde{\hat{G}}^1$, $\varepsilon = 60\%$: Δ reconstructed real component, o reconstructed imaginary component, —— exact real component, ---- exact imaginary component.

Fig. 16. Sample realization solution to model problem III, test function $\tilde{\hat{G}}^1$, $\varepsilon = 100\%$: Δ reconstructed real component, o reconstructed imaginary component, —— exact real component, ---- exact imaginary component.

Fig. 17. Sample realization solution to model problem I, test function $\tilde{\tilde{G}}^2$, $\varepsilon = 20\%$: o reconstructed phase, Δ reconstructed modulus, ---- exact modulus, —— exact phase.

Fig. 18. Sample realization solution to model problem I, test function $\tilde{\tilde{G}}^2$, $\varepsilon = 60\%$: o reconstructed phase, Δ reconstructed modulus, ---- exact modulus, —— exact phase.

Fig. 19. Sample realization solution to model problem I, test function $\tilde{\tilde{G}}^2$, $\varepsilon = 100\%$: o reconstructed phase, Δ reconstructed modulus, ---- exact modulus, —— exact phase.

Fig. 20. Sample realization solution to model problem III, test function $\tilde{\tilde{G}}^2$, $\varepsilon = 20\%$: o reconstructed phase, Δ reconstructed modulus, ---- exact modulus, —— exact phase.

Fig. 21. Sample realization solution to model problem III, test function $\tilde{\tilde{G}}^2$, $\varepsilon = 60\%$: o reconstructed phase, Δ reconstructed modulus, ---- exact modulus, —— exact phase.

Fig. 22. Sample realization solution to model problem III, test function $\tilde{\tilde{G}}^2$, $\varepsilon = 100\%$: o reconstructed phase, Δ reconstructed modulus, ---- exact modulus, —— exact phase.

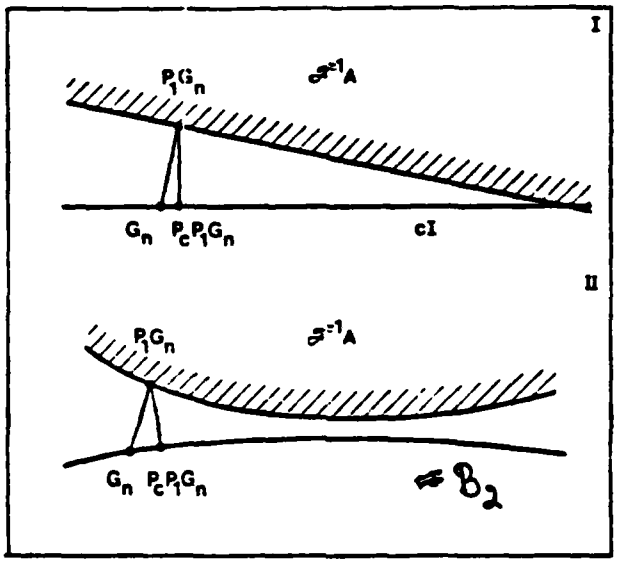

Fig. 1. Two examples of poor convergence of the alternating projection algorithm induced by ill conditioning.

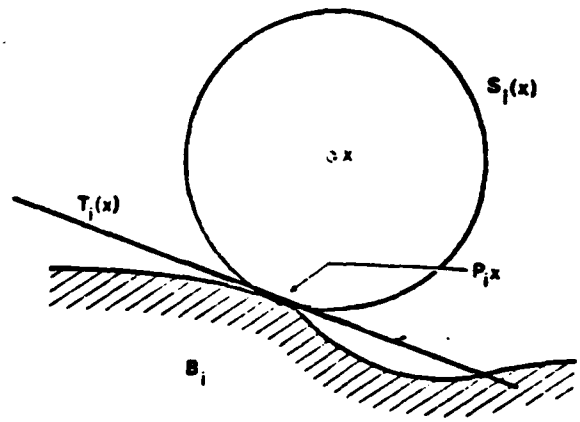

Fig. 2. The tangent plane $T_i(x)$ to a set B_i at the point $P_i(x)$. $T_i(x)$ locally separates B_i and $S_i(x)$.

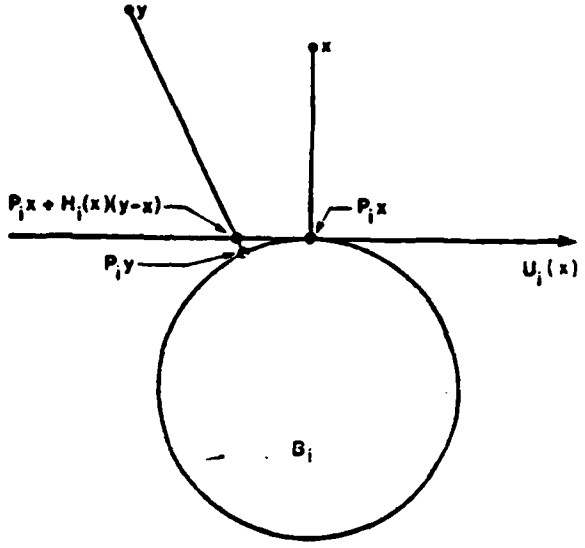

Fig. 3. An example of the form of the Hessian operator $\mathscr{H}_i(x)$ associated with a sphere B_i.

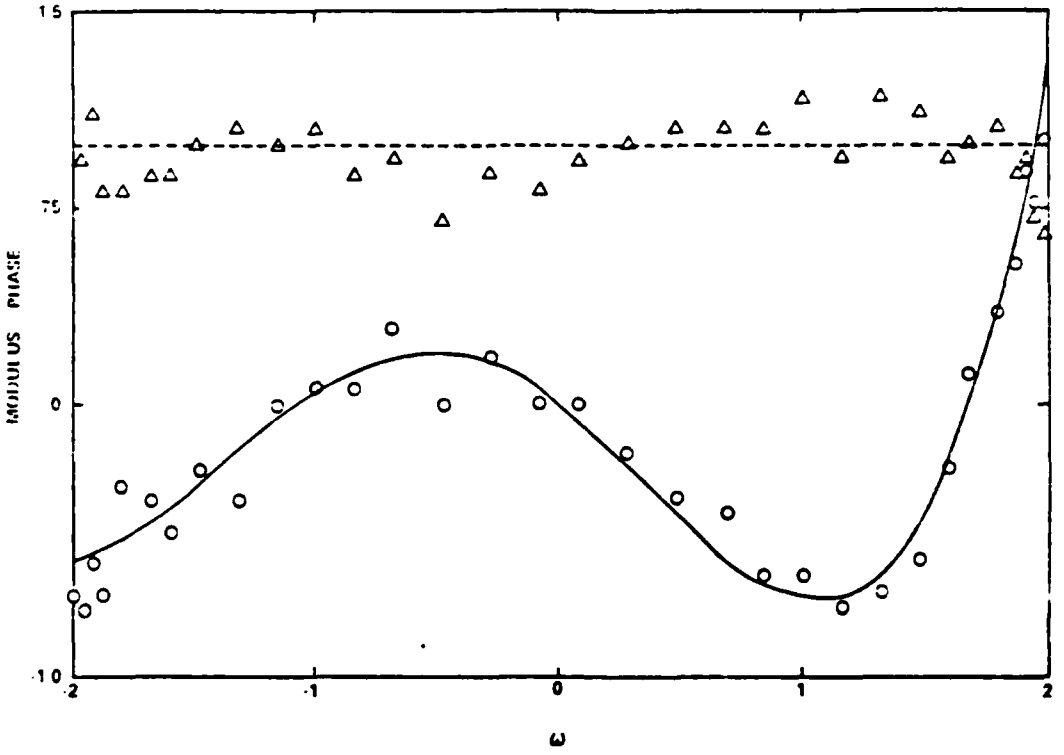

Fig. 4. A solution to model problem I, test function $\tilde{\hat{G}}^2$ with $F(\tilde{\hat{G}}) = 8 \cdot 10^{-5}$: Δ reconstructed modulus, o reconstructed phase, ---- exact modulus, —— exact phase.

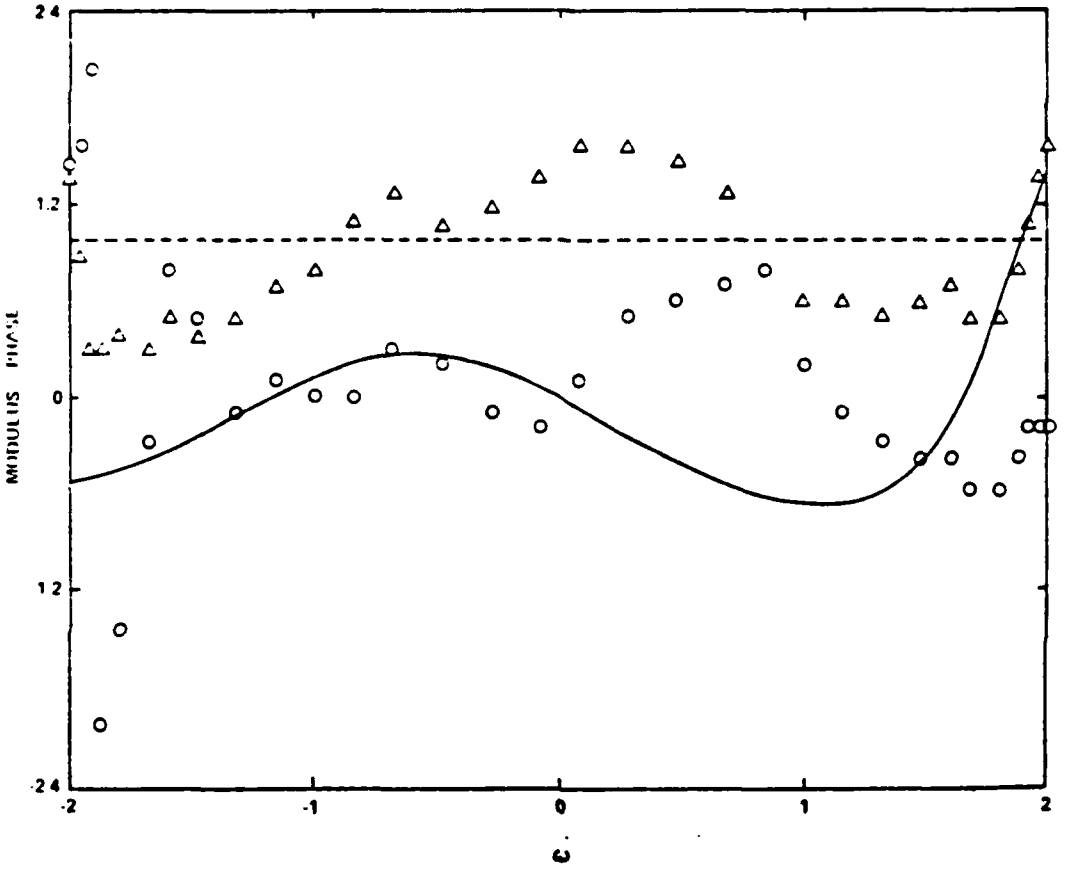

Fig. 5. A solution to model problem I, test function $\tilde{\tilde{G}}^2$ with $F(\tilde{G}) = 7 \cdot 10^{-5}$: Δ reconstructed modulus, O reconstructed phase, ---- exact modulus, —— exact phase.

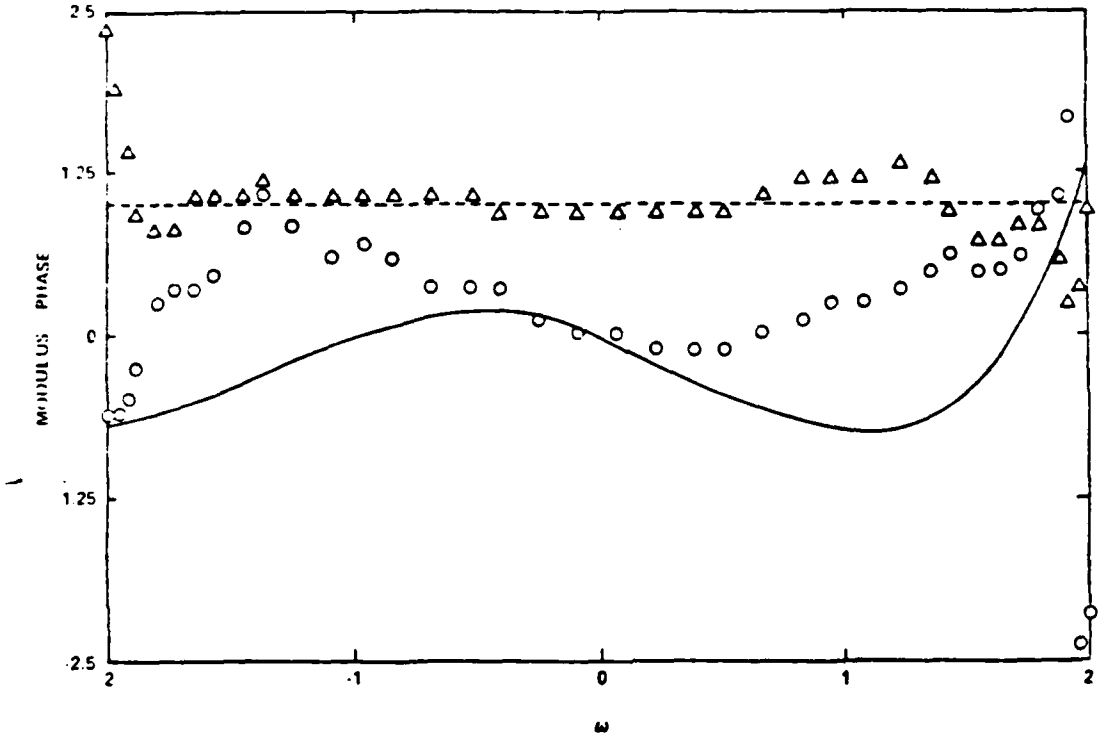

Fig. 6. A solution to model problem I, test function $\tilde{\tilde{G}}^2$ with $F(\tilde{\tilde{G}}) = 8 \cdot 10^{-4}$: Δ reconstructed modulus, o reconstructed phase, ---- exact modulus, —— exact phase.

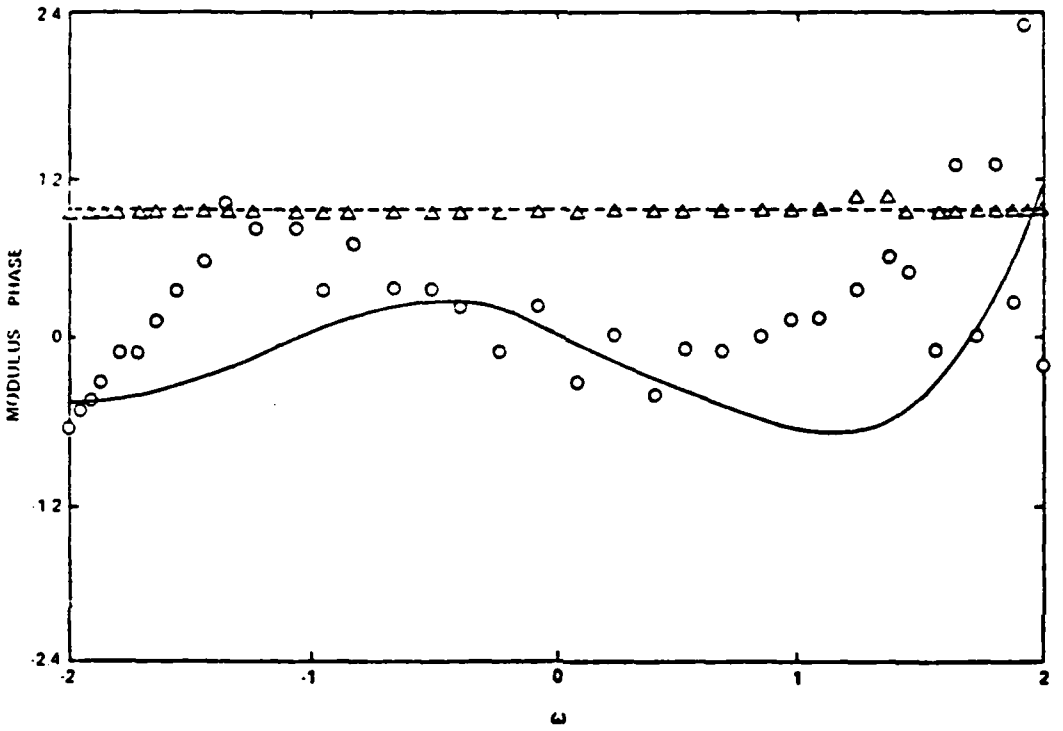

Fig. 7. A solution to model problem III, test function $\overset{\times 2}{G}$ with $F(\tilde{\hat{G}}) = 2 \cdot 10^{-2}$: Δ reconstructed modulus, O reconstructed phase, ---- exact modulus, —— exact phase.

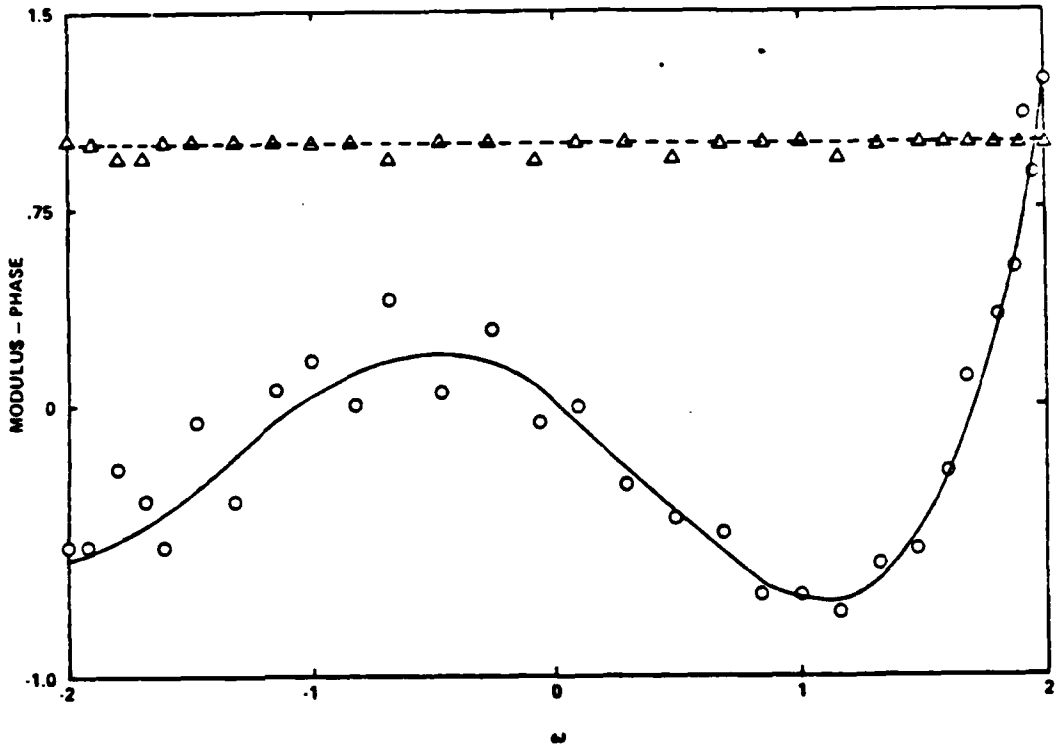

Fig. 8. A solution to model problem III, test function $\tilde{G}^{2}$ with $F(\tilde{G}) = 3 \cdot 10^{-2}$: Δ reconstructed modulus, o reconstructed phase, ---- exact modulus, —— exact phase.

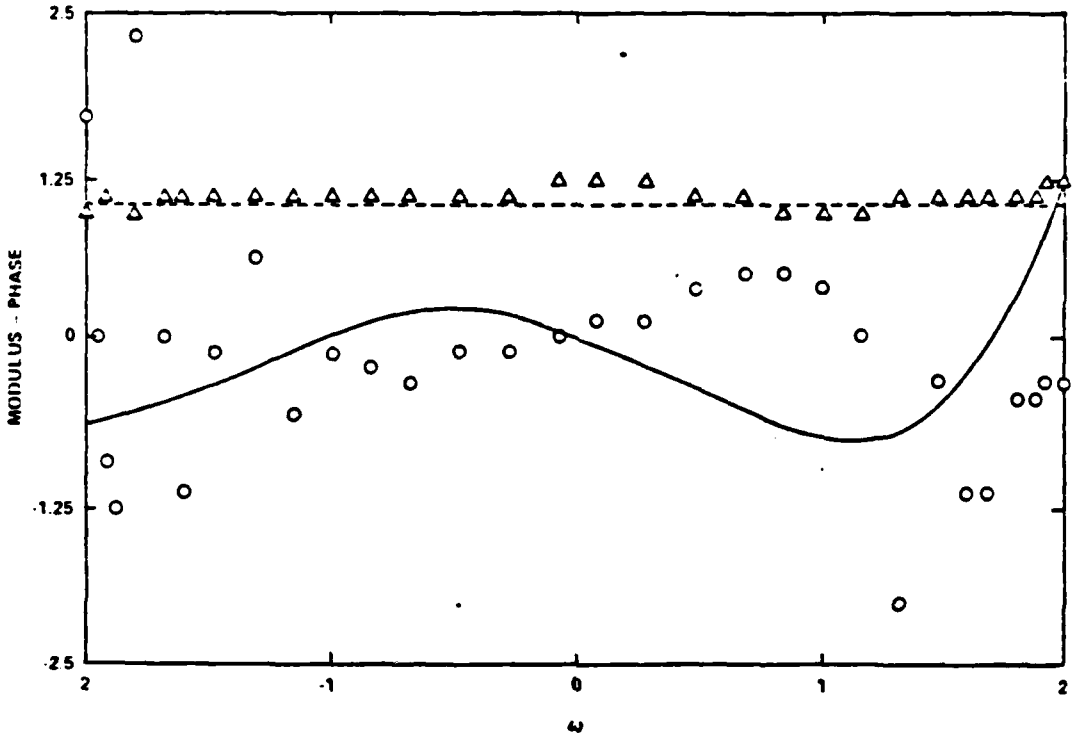

Fig. 9. A solution to model problem III, test function $\tilde{G}^2$ with $F(\tilde{G}) = 7 \cdot 10^{-2}$: Δ reconstructed modulus, o reconstructed phase, ---- exact modulus, —— exact phase.

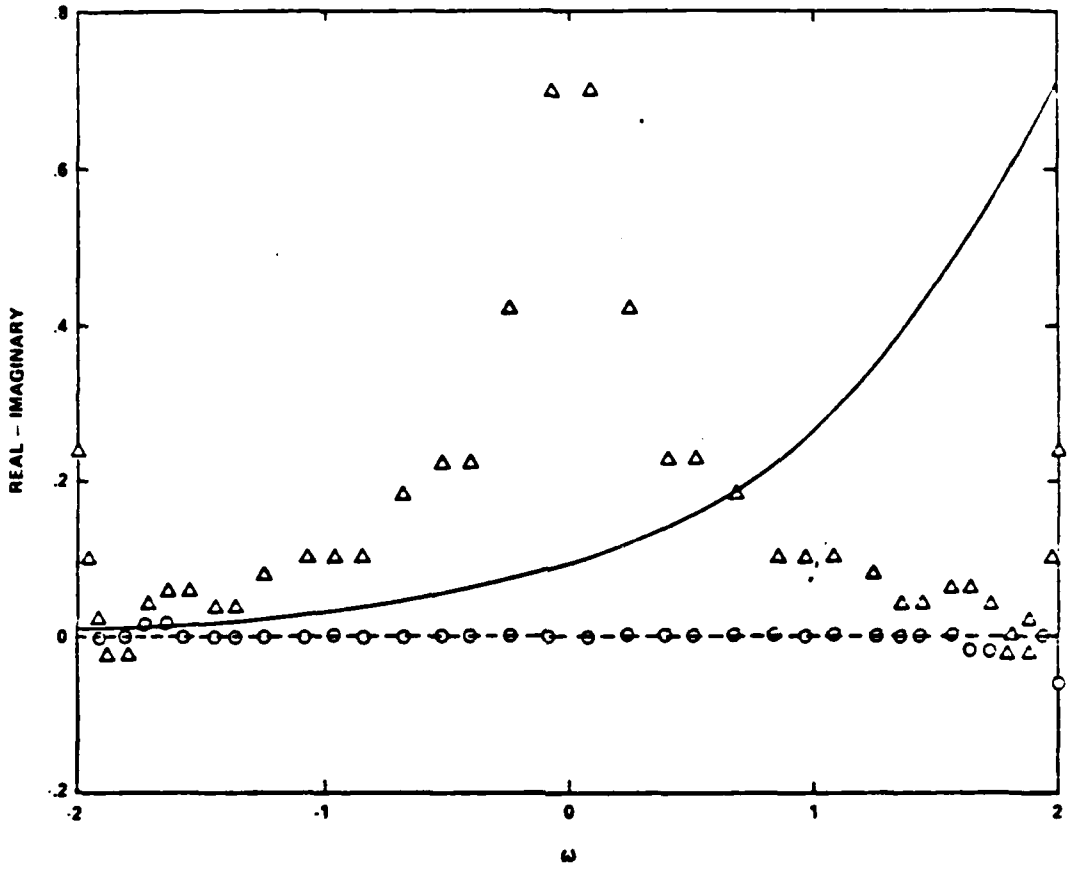

Fig. 10. An example of preservation of symmetry, a solution to model problem I, test function $\tilde{G}^{1}$ with $\hat{G}_0 = 0$.

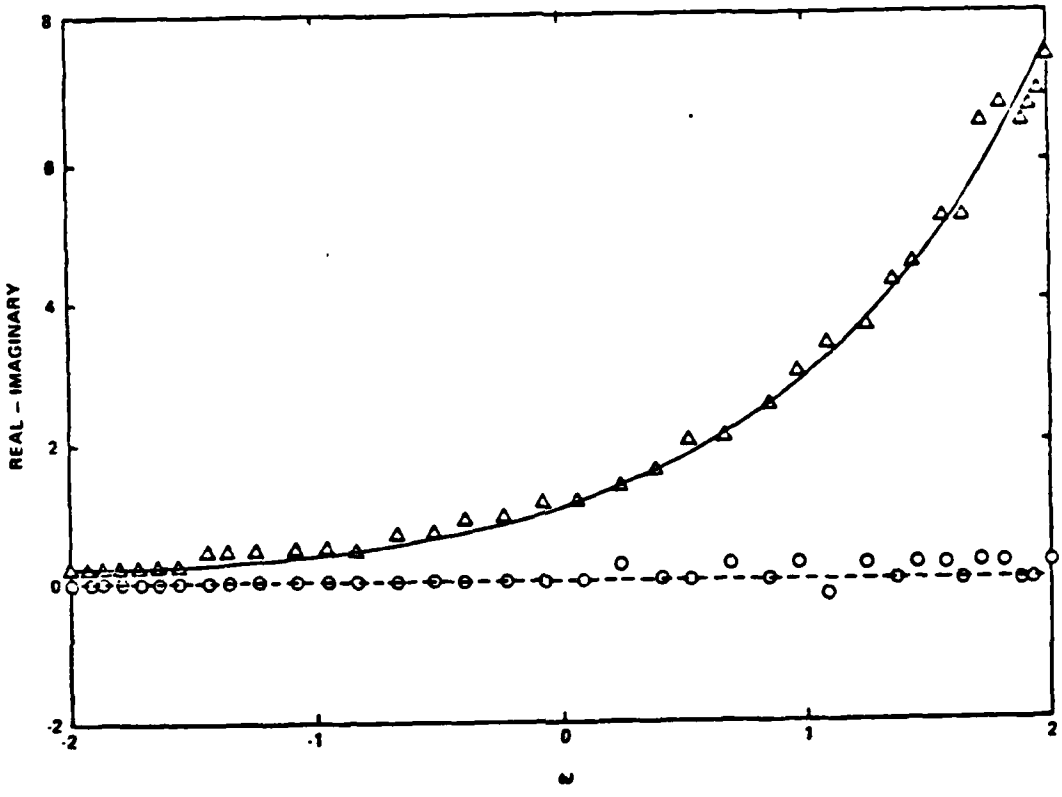

Fig. 11. Sample realization solution to model problem I, test function $\tilde{\hat{G}}^1$, $\varepsilon = 20\%$: Δ reconstructed real component, O reconstructed imaginary component, —— exact real component, ---- exact imaginary component.

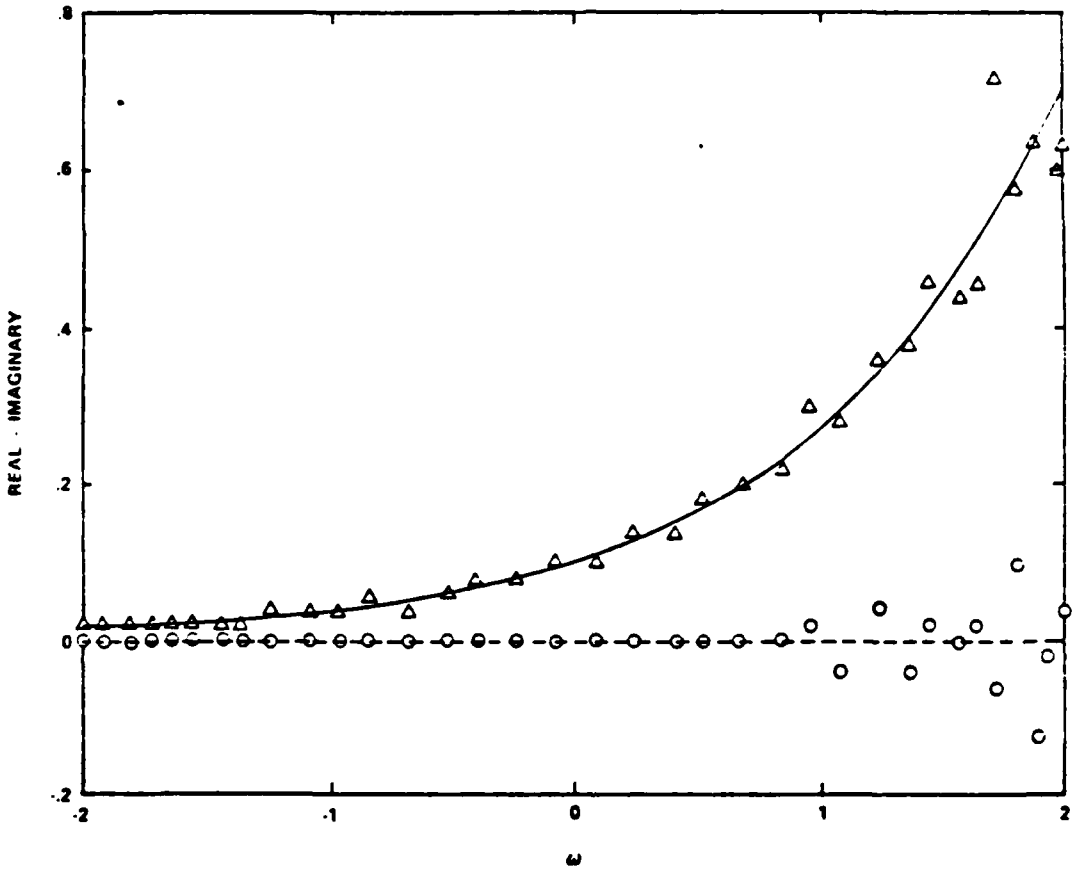

Fig. 12. Sample realization solution to model problem I, test function $\tilde{G}^1$, $\varepsilon = 60\%$: Δ reconstructed real component, $\circ$ reconstructed imaginary component, —— exact real component, ---- exact imaginary component.

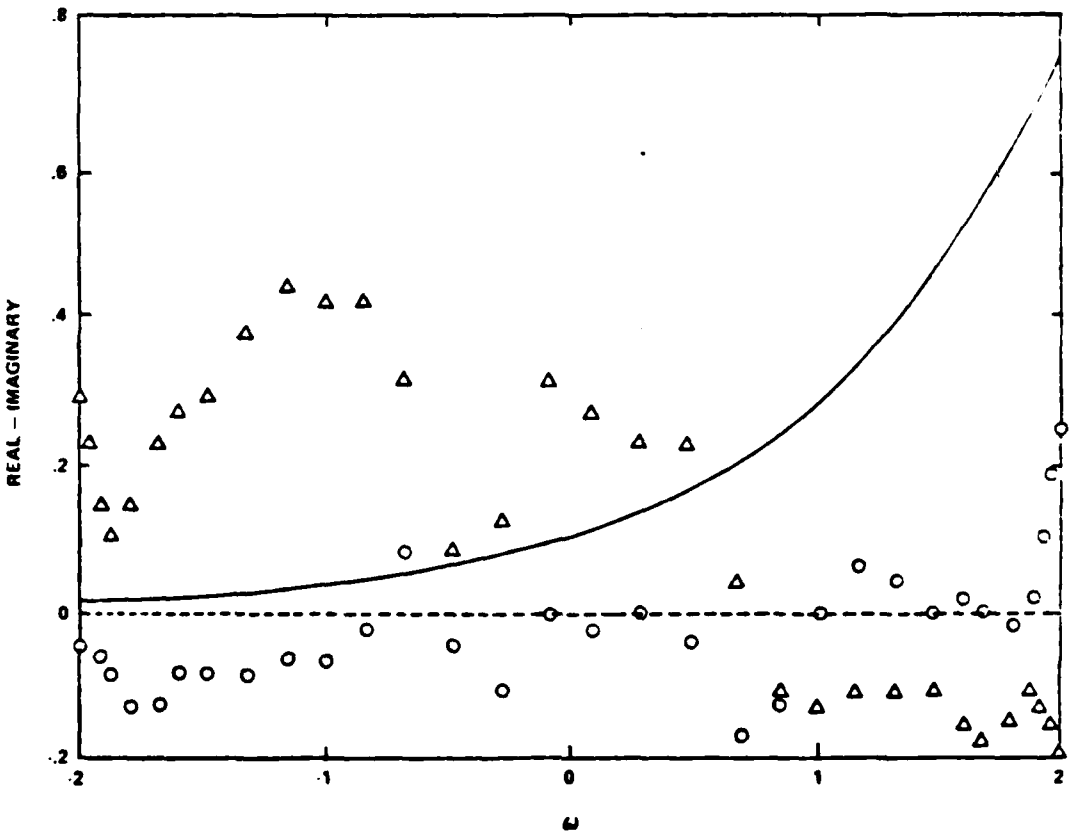

Fig. 13. Sample realization solution to model problem I, test function $\tilde{\tilde{G}}^{1}$, $\varepsilon = 100\%$: Δ reconstructed real component, o reconstructed imaginary component, —— exact real component, ---- exact imaginary component.

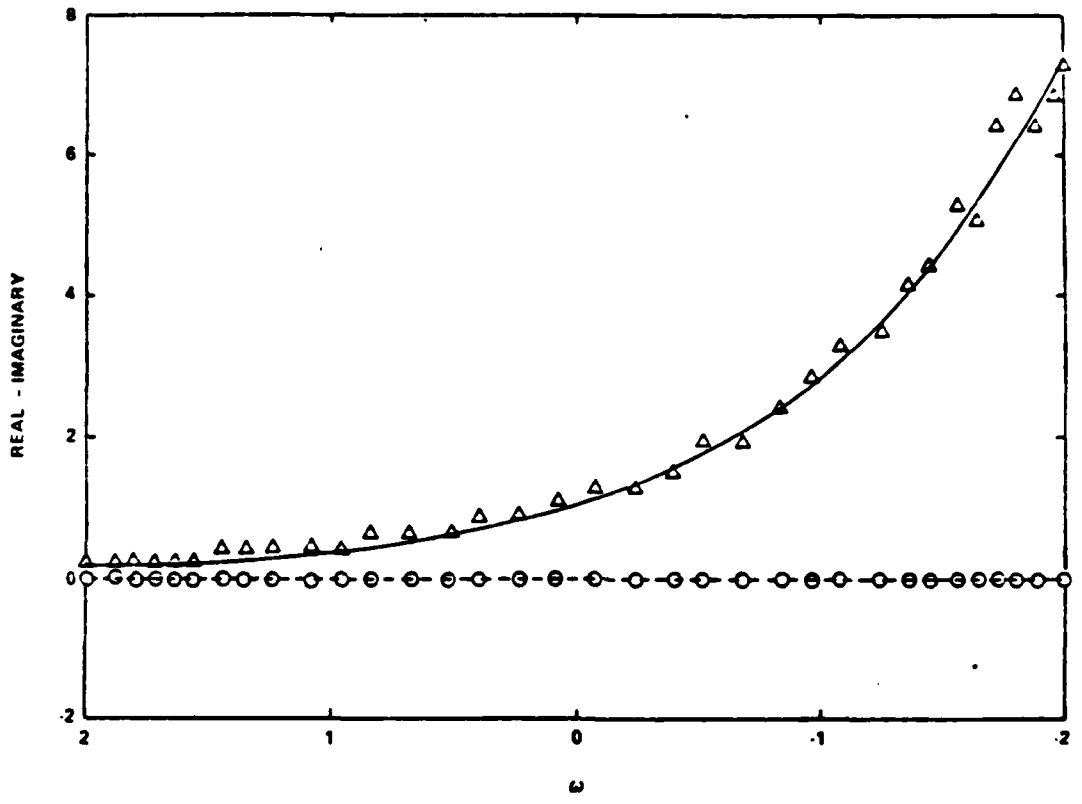

Fig. 14. Sample realization solution to model problem III, test function $\tilde{\hat{G}}^1$, $\varepsilon = 20\%$: Δ reconstructed real component, o reconstructed imaginary component, —— exact real component, ---- exact imaginary component.

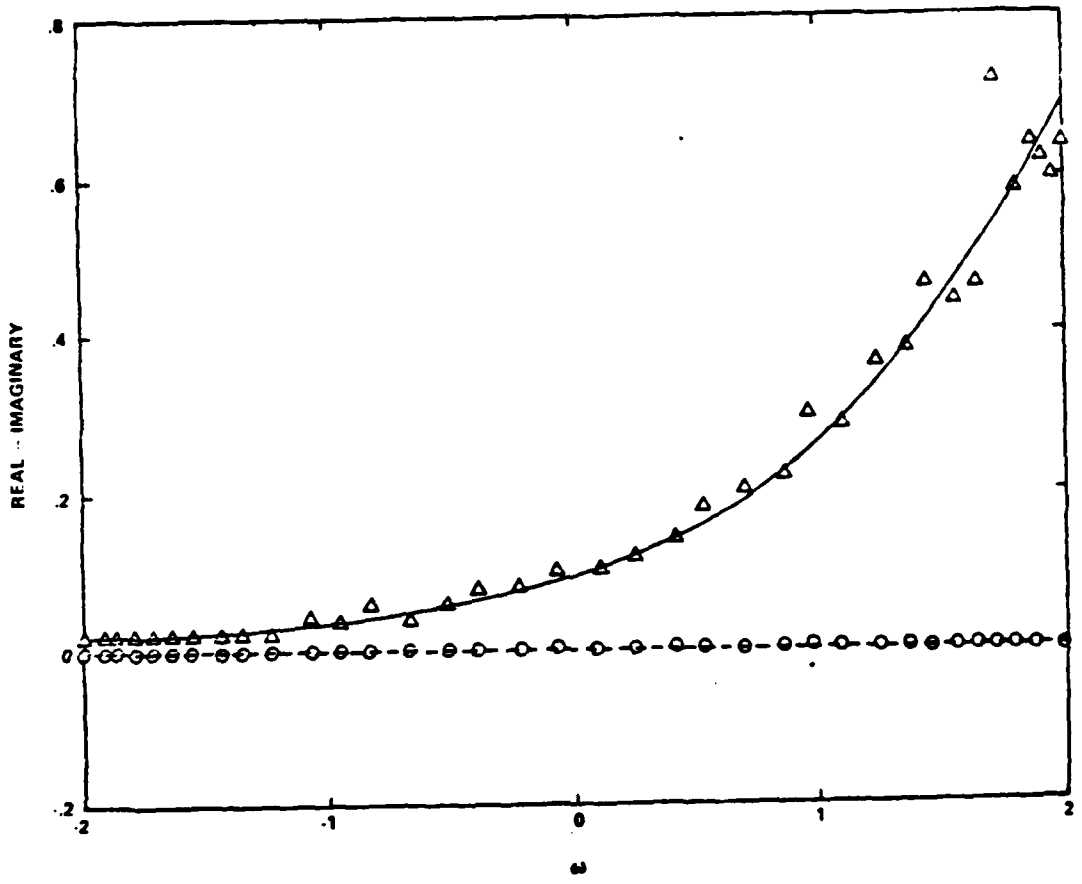

Fig. 15. Sample realization solution to model problem III, test function $\tilde{G}^1$, $\varepsilon = 60\%$: Δ reconstructed real component, o reconstructed imaginary component, —— exact real component, ---- exact imaginary component.

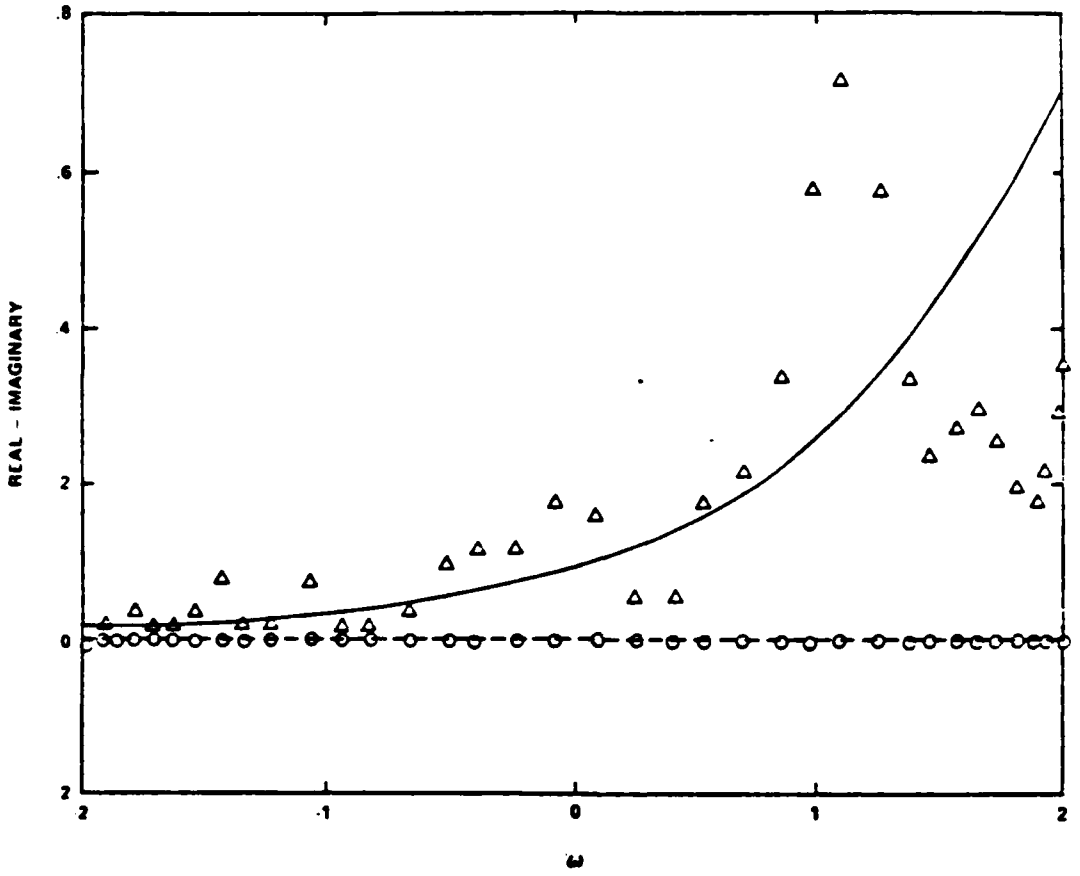

Fig. 16. Sample realization solution to model problem III, test function $\tilde{G}^1$, $\varepsilon = 100\%$: Δ reconstructed real component, O reconstructed imaginary component, —— exact real component, ---- exact imaginary component.

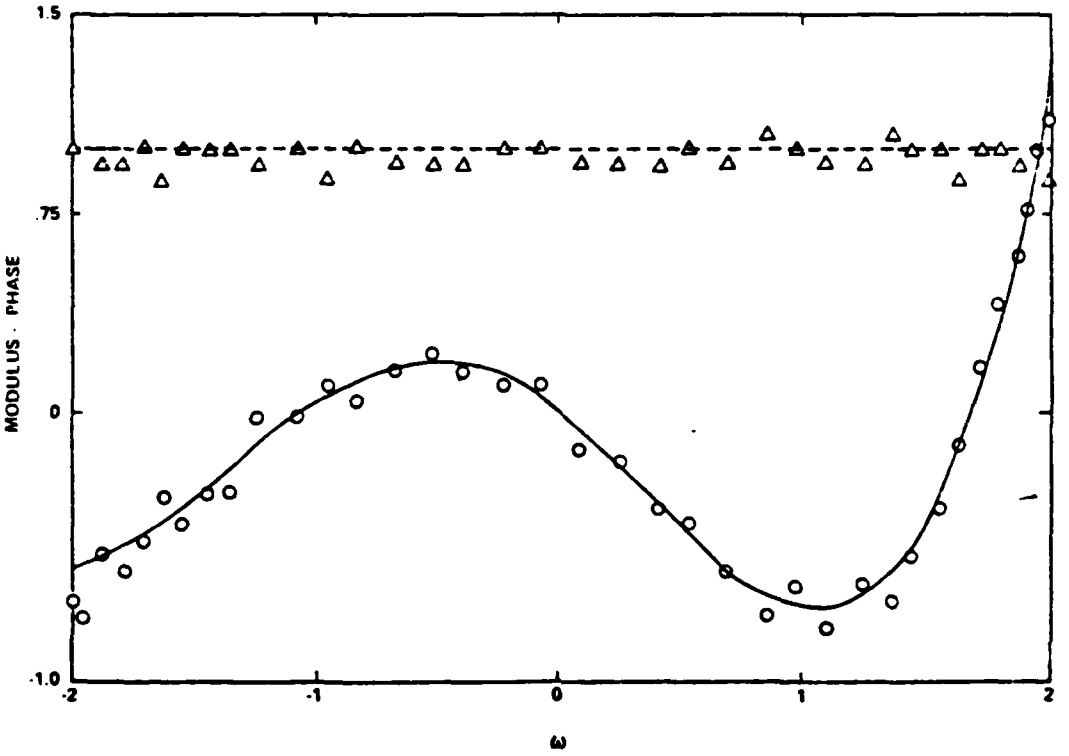

Fig. 17. Sample realization solution to model problem I, test function $\tilde{G}^2$, $\varepsilon = 20\%$: o reconstructed phase, Δ reconstructed modulus, ---- exact modulus, —— exact phase.

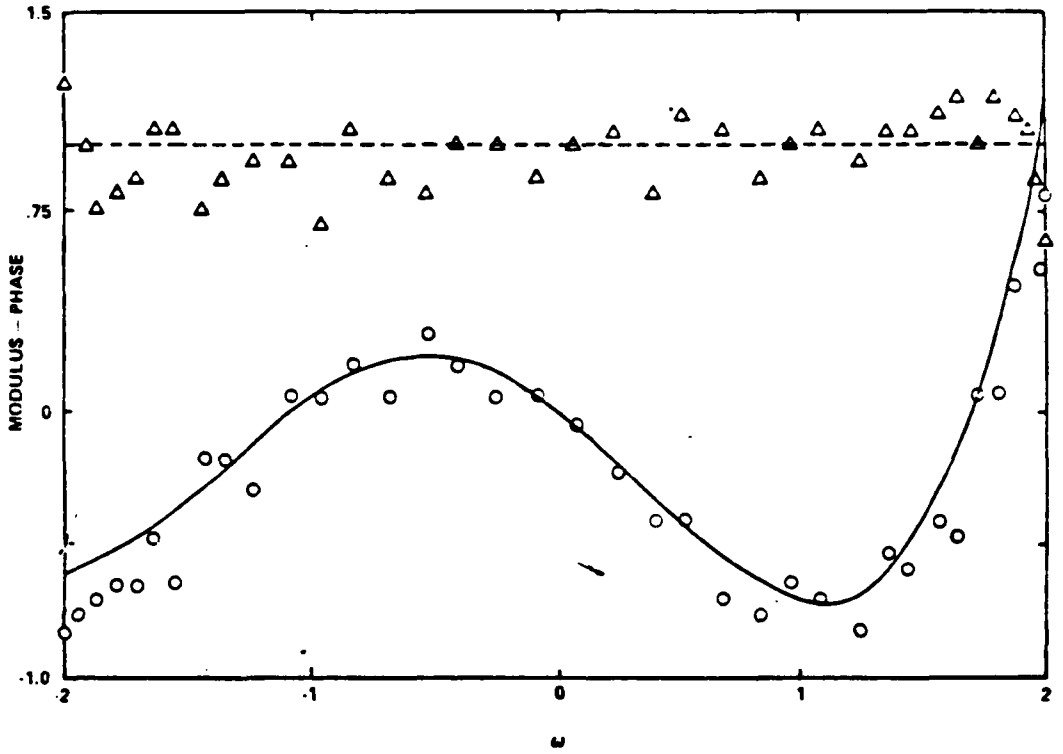

Fig. 18. Sample realization solution to model problem I, test function $\hat{G}^2$, $\epsilon = 60\%$: o reconstructed phase, Δ reconstructed modulus, ---- exact modulus, —— exact phase.

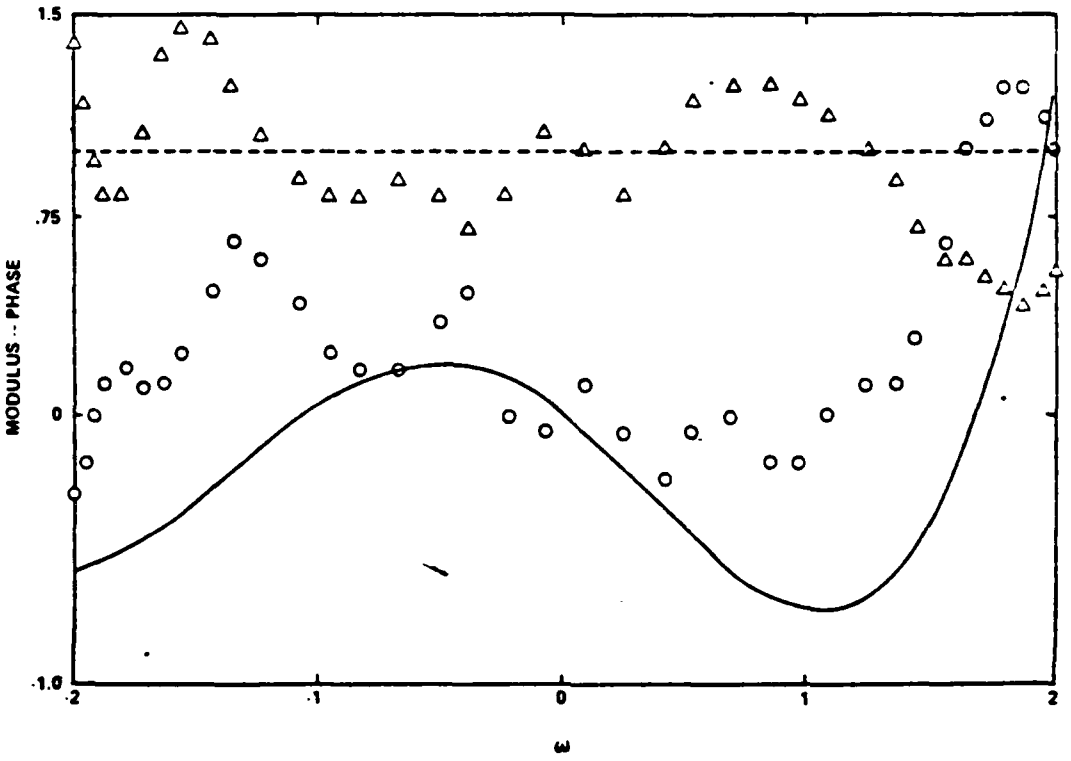

Fig. 19. Sample realization solution to model problem I, test function $\tilde{G}^2$, $\varepsilon = 100\%$: O reconstructed phase, Δ reconstructed modulus, ---- exact modulus, —— exact phase.

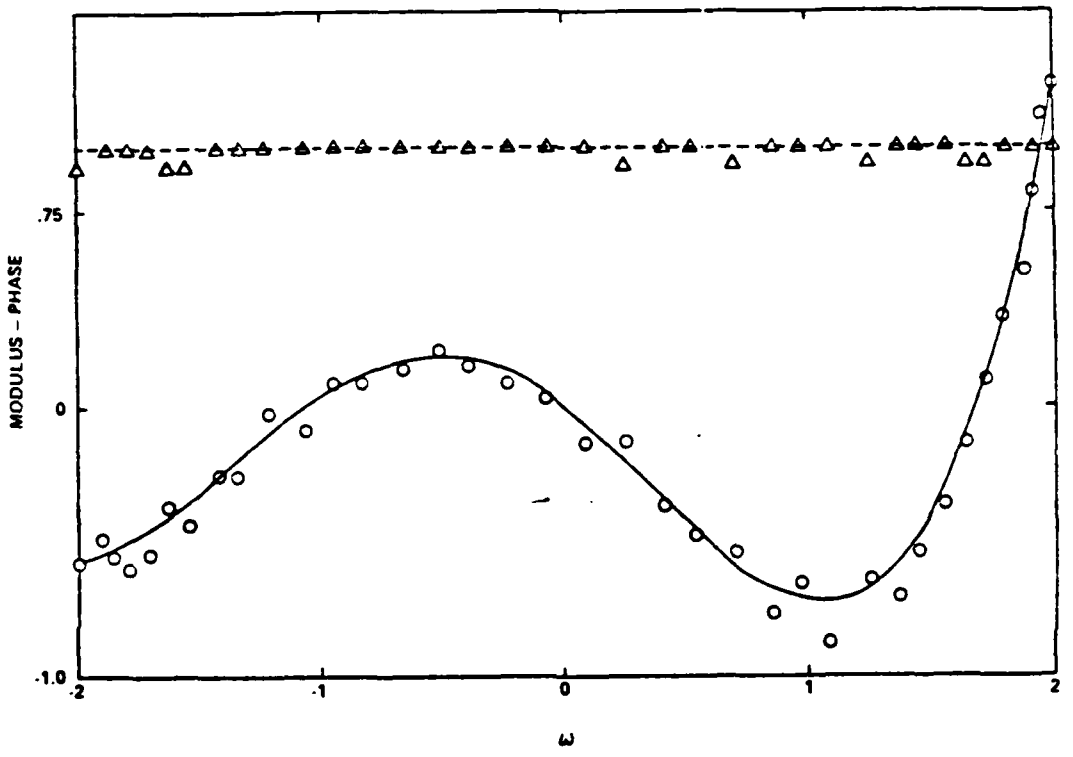

Fig. 20. Sample realization solution to model problem III, test function $\tilde{G}^2$, $\varepsilon = 20\%$: o reconstructed phase, Δ reconstructed modulus, ---- exact modulus, —— exact phase.

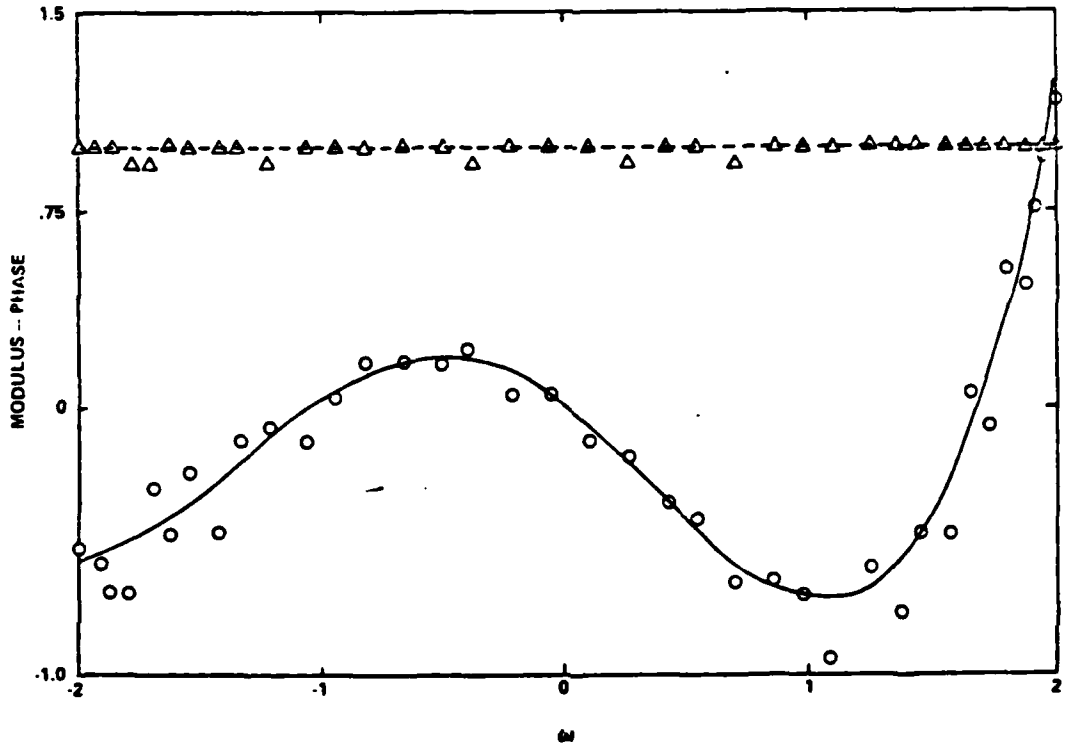

Fig. 21. Sample realization solution to model problem III, test function $\hat{G}^2$, $\varepsilon = 60\%$: o reconstructed phase, Δ reconstructed modulus, ---- exact modulus, —— exact phase.

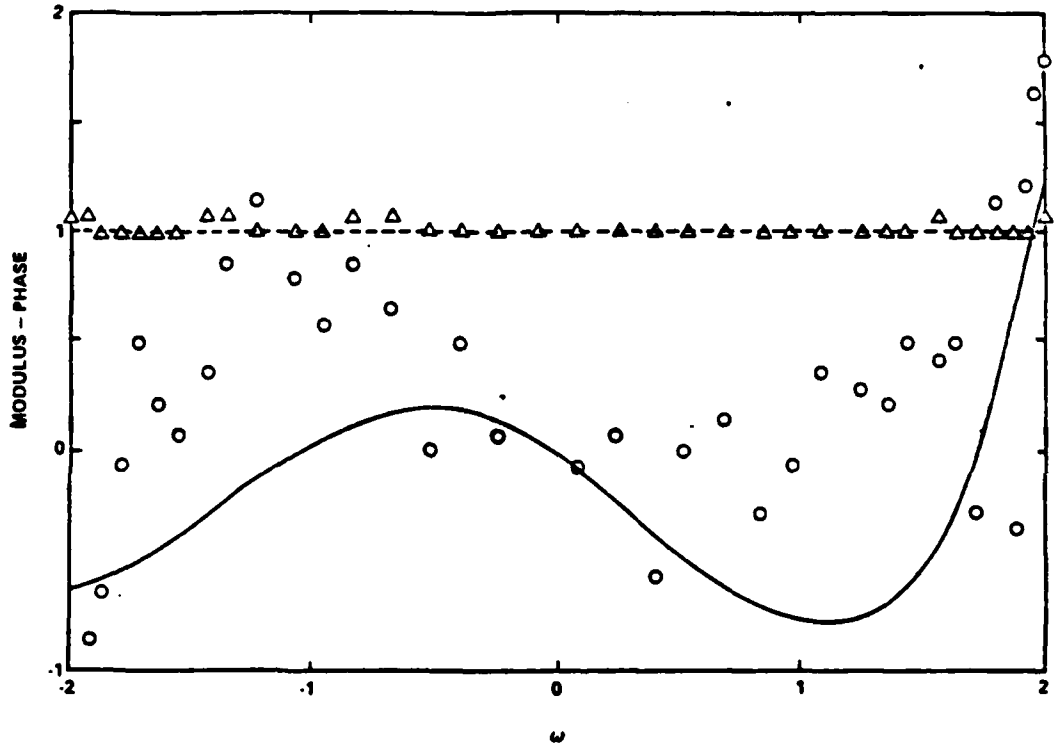

Fig. 22. Sample realization solution to model problem III, test function $\tilde{G}^2$, $\varepsilon = 100\%$: O reconstructed phase, Δ reconstructed modulus, ---- exact modulus, —— exact phase.

Distribution List

addresses	number of copies
Dr RJ Michalak RADC/OCSE	2
RADC/TSTD GRIFFISS AFB NY 13441	1
RADC/DAP GRIFFISS AFB NY 13441	2
ADMINISTRATOR DEF TECH INF CTR DTIC-DDA CAMERON STA BG 5 ALEXANDRIA VA 22314	12
ERIM POBox 8618 Ann Arbor MI 48107	1
ERIM Attn: Dr S Robinson POBox 8618 Ann Arbor MI 48107	1

MRJ Inc. 1
Attn: Dr K Robinson
71 Blake St.
Needham MA 02192

Itek Corp. 1
Attn: E. Galat
Optical Systems Division
10 Maguire Rd.
Lexington MA 02173

DARPA/STO 1
Attn: Lt. Col. Bell
1400 Wilson Blvd.
Arlington VA 22209

AFWL/ARAA 1
Attn: Dr L Skolnik
Kirtland AFB 87117

Aerospace 1
Attn: Dr V Mahajan
MS M4/978
2350 E. El Segundo Blvd.
El Segundo CA 90245

Lockheed 1
Attn: Dr S Williams
3251 Hanover St.
Palo Alto CA 94304

Hughes 1
Attn: Dr R Withrington
Build. E1 MS F/124
El Segundo CA 90245

NOSC
Attn: Dr D Fried
POBox 446
Placentia CA 92670 1

IDM Corp.
Attn: Dr E Silvertooth
C/O Jeanette Miller (Security Officer)
5155 W. Rosecrans Ave Hawthorne CA 90250 1

Riverside Research
Attn: HALO Library, Mr R Passett
1701 N Fort Meyer Dr
Arlington VA 22209 1

DARPA/DEO
Attn: R Strunce
1400 Wilson Blvd.
Arlington VA 22209 1

ERIM
Attn: Dr S Robinson
POBox 8618
Ann Arbor MI 48107 1

MRJ Inc.
Attn: Dr K Robinson
71 Blake St.
Needham MA 02192 1

Itek Corp.
Attn: E. Galat
Optical Systems Division
10 Maguire Rd.
Lexington MA 02173 1

DARPA/STO
Attn: Lt. Col. Bell
1400 Wilson Blvd.
Arlington VA 22209 1

AFWL/ARAA 1
Attn: Dr L Skolnik
Kirtland AFB 87117

Aerospace 1
Attn: Dr V Mahajan
MS M4/978
2350 E. El Segundo Blvd.
El Segundo CA 90245

Lockheed 1
Attn: Dr S Williams
3251 Hanover St.
Palo Alto CA 94304

Hughes 1
Attn: Dr R Withrington
Build. E1 MS F/124
El Segundo CA 90245

tOSC 1
Attn: Dr D Fried
POBox 446
Placentia CA 92670

BDM Corp. 1
Attn: Dr E Silvertooth
C/O Jeanette Miller (Security Officer)
5155 W. Rosecrans Ave Hawthorne CA 90250

Riverside Research 1
Attn: HALO Library, Mr R Passett
1701 N Fort Meyer Dr
Arlington VA 22209

DARPA/DEO 1
Attn: R Strunce
1400 Wilson Blvd.
Arlington VA 22209

RGB Assoc. 1
Attn: Prof R Barakat
POBox 8
Wayland MA 01778

Eikonix Corp. 1
Attn: Dr R Gonsalves
23 Crosby Rd.
Bedford MA 01730

NYIT 1
Institute for Optics
Prof M Halioua
Old Westbury NY 11568

MISSION
of
Rome Air Development Center

RADC plans and executes research, development, test and selected acquisition programs in support of Command, Control Communications and Intelligence (C^3I) activities. Technical and engineering support within areas of technical competence is provided to ESD Program Offices (POs) and other ESD elements. The principal technical mission areas are communications, electromagnetic guidance and control, surveillance of ground and aerospace objects, intelligence data collection and handling, information system technology, ionospheric propagation, solid state sciences, microwave physics and electronic reliability, maintainability and compatibility.